

LUDWIG-MAXIMILIANS-UNIVERSITÄT

THESIS

# Computational methods in Connectomics

*Manuel Berning*

supervised by Dr. Moritz Helmstaedter  
at the Max Planck Institute for Brain Research

September 27, 2017

Erstgutachter: Prof. Alexander Borst  
Zweitgutachter: Prof. Andreas Herz  
Tag der Abgabe: 27.09.2017  
Tag der mündlichen Prüfung: 13.04.2018



## Eidesstattliche Erklärung

Ich versichere hiermit an Eides statt, dass die vorgelegte Dissertation von mir selbständig und ohne unerlaubte Hilfe angefertigt ist.

München, den 14.6.2018

Manuel Berning

## Erklärung

Hiermit erkläre ich, \*

- ☒ dass die Dissertation nicht ganz oder in wesentlichen Teilen einer anderen Prüfungskommission vorgelegt worden ist.
- ☒ dass ich mich anderweitig einer Doktorprüfung ohne Erfolg **nicht** unterzogen habe.
- ☒ dass ich mich mit Erfolg der Doktorprüfung im Hauptfach .....Biologie..... bei der Fakultät für .....Biologie..... der .....Ludwig Maximilians Universität....  
(Hochschule/Universität) unterzogen habe.
- ☐ dass ich ohne Erfolg versucht habe, eine Dissertation einzureichen oder mich der Doktorprüfung zu unterziehen.

München, den 14.06.2018

Manuel Berning

\*) Nichtzutreffendes streichen

# Contents

|           |  |            |
|-----------|--|------------|
| <b>1</b>  | <b>List of Publications</b>  | <b>6</b>   |
| <b>2</b>  | <b>Declaration of contribution as shared first author</b>                  | <b>9</b>   |
| <b>3</b>  | <b>Declaration of contribution as a co-author</b>                          | <b>10</b>  |
| <b>4</b>  | <b>Summary</b>   | <b>11</b>  |
| <b>5</b>  | <b>Introduction</b>  | <b>13</b>  |
| 5.1       | Neurobiology . . . . .   | 13         |
| 5.2       | Structural neurobiology . . . . .  | 17         |
| 5.3       | Connectomics . . . . .   | 20         |
| 5.4       | Electron Microscopy . . . . .  | 23         |
| 5.5       | Barrel cortex . . . . .  | 27         |
| 5.6       | The reconstruction challenge . . . . .                                     | 30         |
| 5.7       | Artificial Intelligence & Machine Learning . . . . .                       | 31         |
| 5.8       | User interaction . . . . .   | 35         |
| <b>6</b>  | <b>Results</b>   | <b>36</b>  |
| 6.1       | SegEM: Efficient image analysis for high-resolution connectomics . . . . . | 36         |
| 6.2       | webKnossos: Efficient online 3D data annotation for connectomics . . . . . | 52         |
| 6.3       | SynEM: Automated synapse detection for connectomics . . . . .              | 61         |
| <b>7</b>  | <b>Summary &amp; Discussion</b>  | <b>126</b> |
| 7.1       | From sparse annotations to volume models . . . . .                         | 126        |
| 7.2       | Novel annotation interactions . . . . .                                    | 130        |
| 7.3       | Automated synapse detection . . . . .                                      | 133        |
| 7.4       | General discussion . . . . .   | 136        |
| 7.5       | Outlook . . . . .  | 137        |
| <b>8</b>  | <b>Bibliography</b>  | <b>140</b> |
| <b>9</b>  | <b>Acknowledgements</b>  | <b>155</b> |
| <b>10</b> | <b>Curriculum Vitae</b>  | <b>156</b> |

# Acronyms

**ATUM** automated tape collecting ultra microtome. 24–26, 133, 134, 136

**B.C.** before Christ. 13

**BOLD** blood oxygen level dependent. 20

**CCD** charge coupled device. 16

**CNN** convolutional neural network. 32, 34, 127, 128

**dTI** diffusion tensor imaging. 20

**EM** electron microscopy. 20, 31, 34, 35, 128–133, 136–138

**fMRI** functional magnetic resonance imaging. 20

**IARPA** Intelligence Advanced Research Projects Activity. 31

**MICRoNS** Machine Intelligence from Cortical Networks. 31

**mSEM** multi-beam scanning electron microscope. 25, 137

**POm** posteromedial complex. 28, 29

**RESCOP** redundant skeleton consensus procedure. 131

**rOTO** reduced osmium thiocarbohydrazide-osmium. 24, 25

**SAC** starburst amacrine cells. 21, 23

**SBEM** serial blockface electron microscopy. 25, 127, 130, 133–136

**SEM** scanning electron microscopy. 24–26, 134, 136

**ssTEM** serial section transmission electron microscopy. 24, 26, 134, 137

**TEM** transmission electron microscopy. 24, 135

**VPM** nucleus ventralis posteromedialis. 28, 29

# 1 List of Publications

**Berning, M., Boergens, K. M., and Helmstaedter, M. (2015). SegEM: efficient image analysis for high-resolution connectomics. *Neuron*, 87(6):1193–1206**

**Abstract** Progress in electron microscopy-based high-resolution connectomics is limited by data analysis throughput. Here, we present SegEM, a toolset for efficient semi-automated analysis of large-scale fully stained 3D-EM datasets for the reconstruction of neuronal circuits. By combining skeleton reconstructions of neurons with automated volume segmentations, SegEM allows the reconstruction of neuronal circuits at a work hour consumption rate of about 100-fold less than manual analysis and about 10-fold less than existing segmentation tools. SegEM provides a robust classifier selection procedure for finding the best automated image classifier for different types of nerve tissue. We applied these methods to a volume of  $44 \times 60 \times 141 \mu m^3$  SBEM data from mouse retina and a volume of  $93 \times 60 \times 93 \mu m^3$  from mouse cortex, and performed exemplary synaptic circuit reconstruction. SegEM resolves the tradeoff between synapse detection and semi-automated reconstruction performance in high-resolution connectomics and makes efficient circuit reconstruction in fully-stained EM datasets a ready-to-use technique for neuroscience.

**Contributions** All work was performed by M.B. and M.H.; K.M.B. acquired the cortex dataset.

**Copyright** Reprinted from [Berning et al., 2015], Copyright 2015, with permission from Elsevier.

**Boergens, K. M., Berning, M., Bocklisch, T., Bräunlein, D., Drawitsch, F., Frohnhofen, J., Herold, T., Otto, P., Rzepka, N., Werkmeister, T., Werner, D., Wiese, G., Wissler, H., and Helmstaedter, M. (2017). webknossos: efficient online 3d data annotation for connectomics. *Nature Methods*, 14:691–694**

**Abstract** We report webKnossos, an in-browser annotation tool for 3D electron microscopic data. webKnossos provides flight mode, a single-view egocentric reconstruction method enabling trained annotator crowds to reconstruct at a speed of  $1.5 \pm 0.6$  mm/h for axons and  $2.1 \pm 0.9$  mm/h for dendrites in 3D electron microscopic data from mammalian cortex. webKnossos accelerates neurite reconstruction for connectomics by 4- to 13-fold compared with current state-of-the-art tools, thus extending the range of connectomes that can realistically be mapped in the future.

**Contributions** M.H. initiated and supervised the project; K.M.B., M.B., T.B., N.R., T.W. and M.H. developed specifications and conceptual design with contributions by H.W.; T.B., D.B., J.F., T.H., P.O., N.R., T.W., D.W., G.W. and K.M.B. implemented the software; H.W., M.B., K.M.B. and F.D. provided data; K.M.B., M.H., H.W. and M.B. analyzed the data; M.H., K.M.B. and M.B. wrote the manuscript with contributions by all authors.

**Copyright** Reprinted by permission from Macmillan Publishers Ltd: Nature methods [Boergens et al., 2017], copyright 2017.

**Staffler, B., Berning, M., Boergens, K. M., Gour, A.,  
van der Smagt, P., and Helmstaedter, M. (2017).**

**SynEM: Automated synapse detection for connectomics.  
eLife, 6:e26414**

**Abstract** Nerve tissue contains a high density of chemical synapses, about 1 per  $\mu m^3$  in the mammalian cerebral cortex. Thus, even for small blocks of nerve tissue, dense connectomic mapping requires the identification of millions to billions of synapses. While the focus of connectomic data analysis has been on neurite reconstruction, synapse detection becomes limiting when datasets grow in size and dense mapping is required. Here, we report SynEM, a method for automated detection of synapses from conventionally en-bloc stained 3D electron microscopy image stacks. The approach is based on a segmentation of the image data and focuses on classifying borders between neuronal processes as synaptic or non-synaptic. SynEM yields 97% precision and recall in binary cortical connectomes with no user interaction. It scales to large volumes of cortical neuropil, plausibly even whole-brain datasets. SynEM removes the burden of manual synapse annotation for large densely mapped connectomes.

**Contributions** Conceived and initiated the project: MH; supervised the project: MH and PvdS; Developed algorithms, implemented algorithms, analyzed data: BS; provided segmentations and contributed to algorithm development: MB; provided EM data: KMB; provided expert synapse annotations: AG; wrote the paper: MH and BS.

**Copyright** This article [Staffler et al., 2017] is licensed under the CC BY 4.0.

**Note** Accepted manuscript, PDF only. Full online edition to follow.

## 2 Declaration of contribution as shared first author

The paper [Boergens et al., 2017] reports the development, implementation and testing of an efficient 3D data viewing and annotation tool for connectomics.

Key components of this paper are: the conceptual design of the tool; the discovery of flight mode as an intuitive 3D data interaction in EM data; the extensive testing of annotation modes and the associated effects on tracing speed and accuracy; the software implementation; comparison to state-of-the-art tools; implementation of the connectome-reconstruction workflow.

The two first authors (KMB and MB) have contributed to these as follows: MB contributed to the conceptual design of the tool; KMB was instrumental in the discovery of flight mode as an intuitive data mode; MB and KMB contributed to extensive testing of annotation modes in an about 60% / 40% share; software implementation was performed by the scalable minds team; the comparison to other tools was performed primarily by MB with contributions by KMB; the connectome reconstruction workflow and its results were contributed by KMB.

---

Kevin Michael Boergens

---

Manuel Berning

### 3 Declaration of contribution as a co-author

The paper [Staffler et al., 2017] describes the development, implementation and evaluation of an automated method for synapse detection in 3D electron microscopy data.

Key components of this paper are: the conceptual design of the method, the implementation of the feature extraction pipeline for interfaces, the formulation and implementation of the classification task, the single synapse performance measurements, the synapse ground truth generation and the model for connectome error based on single synapse performance measurements.

MB has contributed: Provided segmentations of the main dataset used for development and evaluation of the method, contributed to developing the approach of synapse classification as an interface classification task, wrote initial implementations of the interface detection, texture features and pooling statistics. Contributed part of the tracings used in the sparse local connectome example.

---

Moritz Helmstaedter

---

Manuel Berning



## 4 Summary

Understanding the brain remains one of the frontiers of scientific discovery. While many biophysical mechanisms in the brain have been studied and the stimuli that single neurons or brain areas are responsive to have been mapped, the scientific community has not yet been able to reproduce some of its most exciting capabilities or to simulate its behavior. While not sufficient on their own, cellular resolution maps of connectivity and the measurement of their variability and conserved properties between animals, species and brain regions will be of great importance for understanding the processing in the brain. This thesis focuses on the challenge of mapping neuronal connections within a piece of brain tissue with single cell and single synapse resolution using electron microscopy.

The progress towards a complete description of wiring is limited by two main aspects: The first one is data acquisition with current techniques requiring a tradeoff between imaging throughput and quality of the resulting dataset. The quality has to be sufficient to reconstruct all neurites and identify all synapses contained in a given tissue block. The other is the manual annotation effort required to extract the neuronal connectivity graph from such a dataset. Both of these challenges have been addressed in the last decade by partial automation of data acquisition and analysis. Data acquisition using a Serial Block Face Scanning Electron Microscope chooses a trade-off between data acquisition speed and quality which enables mapping of synaptic circuits in mammals today. Here I present methodological improvements for data analysis in such datasets.

In this thesis three projects are reported. The first project describes a method for generating volume models of neurons from sparse annotations in conventionally stained electron microscopy data by using machine learning for automated inference of membrane position and subsequent segmentation by a seeded watershed algorithm. This method of reconstructing neurites in such a dataset requires about 100-fold less manual annotation time than manual volume segmentation and about 10-fold less than other reconstruction methods published at that time.

In the second project a browser based display and annotation tool for 3D electron

microscopy datasets was developed and the effects of novel features of this tool on reconstruction speed and accuracy were quantified. This tool called webKnossos provides capabilities for efficient generation of annotations by many collaborators in parallel and its advantages over existing tools, especially focused on data delivery and a novel annotation mode, were quantified. The tool speeds up data annotation by 4-13 fold.

In the third project an automated method for detecting chemical synapses in such datasets is proposed and evaluated. The approach classifies all borders between segments as either synaptic or non-synaptic based on feature calculations on different sub-volumes close to the interface area of the two involved segments. This methods can generate binary connectomes with above 97% precision and recall without any manual annotation.

Together these projects provide substantial gains in efficiency for connectomic data analysis including a significant increase in efficiency of manual annotation for wire reconstruction and a method for complete automation of synapse detection.

# 5 Introduction

I regard the brain as a computer which will stop working when its components fail. There is no heaven or afterlife for broken down computers; that is a fairy story for people afraid of the dark.

---

Stephen Hawking

## 5.1 Neurobiology

While most scientists would agree with the quote above, scientists on the other hand do not yet have mechanistic explanations for the properties of the human brain like for computers and their components. One property that is currently only attributed to the human brain is the emergence of consciousness, which is defined as "the quality or state of being aware especially of something within oneself" [Merriam-Webster, 2017].

This special property of the brain has fascinated humans for a long time with first documented cases of cranial trephining, the opening of the skull with primitive tools, dated back to the Neolithic Age (8000 to 3000 before Christ (B.C.)) [Prioreschi, 1991]. The author suggests that these surgeries were performed with the goal of "recall of dead individuals to life", which today would rather be described as a temporary loss of consciousness. While this procedure does not imply an understanding of the brain as the seat of consciousness or the concept of consciousness itself, it can be seen as an indication of knowledge about the correlation of head trauma with the loss of consciousness.

Hippocrates stated around 400 B.C.: "Men ought to know that from the brain, and from the brain only, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear, and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant" [Jones et al., 1952]. This statement still sums up the reasons for the large interest in understanding this organ and its emergent properties today.

Research on the brain with modern scientific methods has started to make significant progress over the last few centuries. A first indication that electrical current is involved in muscle movement was discovered by Luigi Galvani, who observed muscle movement in frog legs when he created an electronic circuit by connecting a steel wire with a brass wire and attaching both ends to specific locations of a frog [Galvani and Aldini, 1792]. Galvani did not yet understand that he had created an electronic circuit by exposing the two different metals to the salt water in the frog which acted as an electrolyte and rather thought that the frog generated the "animal electricity" itself. This led to a controversy with Alessandro Volta who shortly afterwards showed how metals of different kind in an electrolyte can generate electrical current and thereby invented the first battery [Volta, 1800].

Experimental evidence that certain functions are localized in the brain of pigeons was provided by behavioral research coupled with lesions carried out by Marie Jean Pierre Flourens. These experiments provided indications of the localization of motor coordination in the cerebellum, perceptions and judgments in the cerebral hemispheres and at least some vital functions in the medulla oblongata [Flourens, 1824].

In the mid 19th century work by Paul Pierre Broca and Carl Wernicke showed that higher human mental abilities are also localized in the brain by studying patients with lesions in the brain. Broca showed that problems in speech synthesis are correlated with lesions in a region in the posterior part of the frontal lobe in the left hemisphere [Broca, 1865], while Wernicke showed a similar correlation of speech comprehension with lesions in the posterior part of the temporal lobe in the left hemisphere [Wernicke, 1874].

At about the same time Fritsch and Hitzig managed to show that exciting the cerebrum in certain locations with electrical currents will selectively make muscles of dogs contract [Fritsch and Hitzig, 1863]. This combined the insight about the role of electricity in our nervous system with the localization of brain function in a single experiment and thus summarizes the main discoveries about the brain at this point in time.

The next important insight was the neuron doctrine, stating that the brain is made up of many clearly delineated building blocks which are the main signaling units in the brain. The doctrine was formulated by Santiago Ramon y Cajal [Cajal, 1888] after using the staining procedure developed by Camillo Golgi [Golgi, 1873] to analyze the structure of sparse collections of neurons within the brain, see Figure 5.1 for an example. Thus at the beginning of the 20<sup>th</sup> century two important features about the structure of the brain had been studied, the localization of certain functions of the brain to specific regions and that each region is made up of many delineated building blocks.

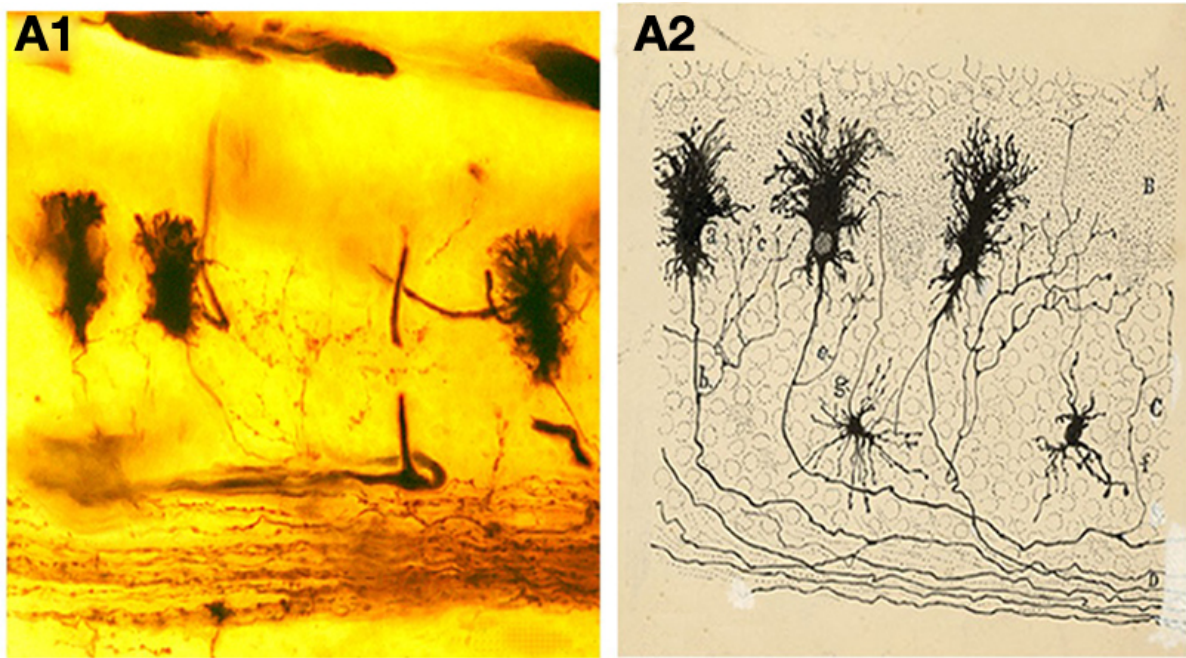


Figure 5.1: **Golgi stain (A1) and related drawing (A2)** of embryonic purkinje cells from a newborn dog by Ramon Cajal. Reproduced from [Garcia-Lopez et al., 2010].

The role of electricity in the mechanisms of the nervous system has been studied in much detail in the 20<sup>th</sup> century since Galvani's initial discovery. For example extracellular microelectrode recordings in the cat cortex and hippocampus were performed and the observed electrical phenomena were described including the effects of different agents used for anesthesia and "rapid deflections of about 1ms in duration" [Renshaw et al., 1940], which would today be called "action potentials". Initial experiments with respect to the biophysics of the cell membrane focused on the squid giant axon due to its large diameter, which made it more accessible to early experimental techniques like conductance and capacity measurements of the membrane during activity [Cole and Curtis, 1939] and initial current and voltage clamp experiments [Marmont, 1949]. A biological insight from this work is the measurement [Hodgkin et al., 1952] and mechanistic explanation [Hodgkin and Huxley, 1952] of action potential transmission based on the properties of sodium and potassium ion channels kinetics.

Patch clamp recordings have even made it possible to measure the electric potential across a patch of (cell) membrane with high temporal accuracy, which allows to measure the properties of single ionic channels in the membrane [Neher and Sakmann, 1976,

Sakmann and Neher, 1984] in denervated frog muscle fibers or slice preparations of the tissue. Experiments using this technique have also led to a good understanding of the kinetics of other ion channel types, e.g. the family of calcium ion channels [Tsien, 1983]. Using a slight variation of the technique, called whole cell patch clamp, one can measure or influence the membrane potential of cells as well as use dyes to stain the recorded cells afterwards [Markram et al., 1997b]. Furthermore this technique has since been extended to allow recording sub-threshold cell membrane potential in response to presented stimuli in numerous brain regions of interest in awake mice [Margrie et al., 2002]. The limitation of these approaches is the small number of cells (or patches) that can routinely be recorded in parallel or in the same sample or animal in general.

An approach for recording the activity of larger assemblies of neurons over longer periods of time is measuring their intracellular calcium levels (see [Grienberger and Konnerth, 2012] for a review of the development of calcium imaging) using either chemical, based on the BAPTA molecule, or genetically encoded calcium indicators, like Cameleons [Miyawaki et al., 1997] or GCaMP [Nakai et al., 2001]. These intracellular calcium sensors change their fluorescent properties based on intracellular calcium levels and their fluorescent responses can then be viewed with a fluorescence microscope and recorded with a charge coupled device (CCD). In order to achieve higher resolutions other microscopes, for example confocal [Minsky, 1961], two photon [Denk et al., 1990] or stimulated emission depletion microscopes [Hell and Wichmann, 1994] can be used. All these microscopy methods have different advantages, usually trading off speed, field of view, tissue requirements and resolution of the imaging procedure. These developments have generated information about everything from sub-cellular response properties [Euler et al., 2002] to the activity of almost all cells of certain species measured simultaneously [Ahrens et al., 2013] while presenting different stimuli or perturbing the system using pharmacological agents or optogenetics.

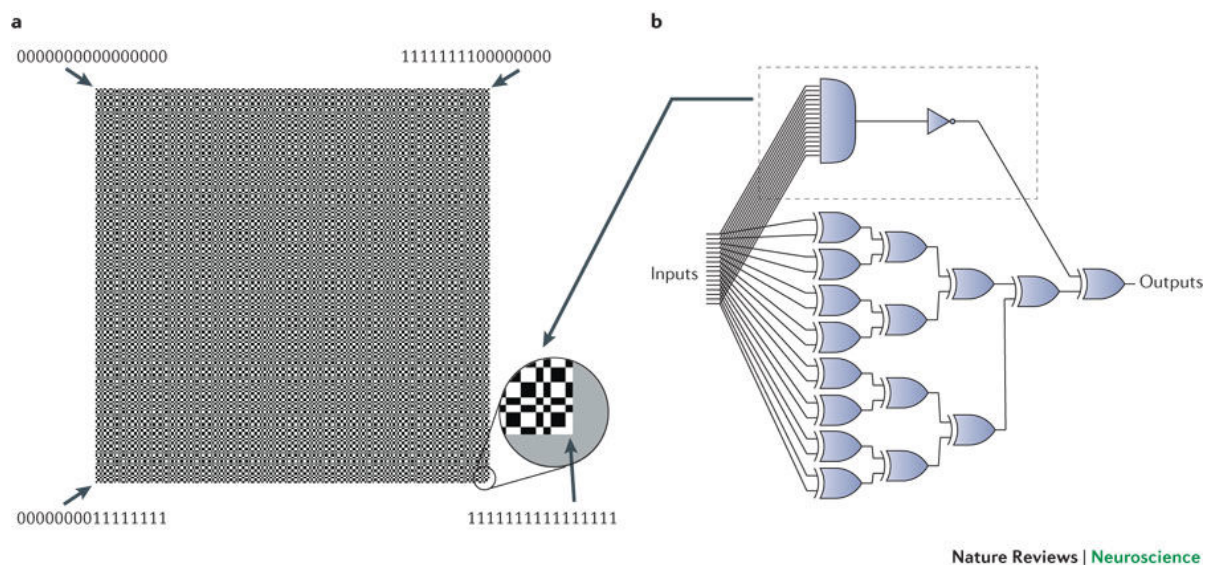
These developments have given valuable insights into the correlations of activity in many neurons types from a diverse set of brain regions with the presentation of stimuli and the biophysical mechanisms used during these computations. A common practical challenge is determining relevant configurations in the input space for a given neuron in a higher brain area due to the large number of inputs to the brain and its inherent state.

## 5.2 Structural neurobiology

Structural biology is concerned with the spatial arrangement of atoms within biological molecules and how these provide a certain functionality. For example how does the spatial arrangement of atoms in the molecule forming an ion channel in a cell membrane make it selective for certain ions [Doyle et al., 1998, Jiang et al., 2003], or how proteins involved in the communication between the individual neurons at synaptic junctions [Südhof, 1995, Südhof and Rothman, 2009] make the signal transmission fast and specific. These results have generated mechanistic explanations of the selectivity of ion channels, the molecular machinery of synapses and transcription and translation of genes.

In analogy the field of structural **neurobiology** can be defined as describing the structure at another level of abstraction: While structural biology describes the spatial arrangement of atoms within molecules, structural neurobiology describes the structure of cells and synapses within the information processing network of the brain. Arguments have been made that structural neurobiology is an important missing link in the attempt to understand the behavior of the brain and the algorithms it implements [Denk et al., 2012]. The analogy to a computational device in this paper makes an important point: There are systems in which mapping the entirety of the functional responses to all possible inputs might yield a much more complex and and less efficient way of representation than the structural analysis of the underlying circuits, see Figure 5.2. Furthermore the argument that knowledge of the connectivity will at least help with the selection between competing models of neural computation is presented. The analogy of the brain to an electronic circuit does have its limitations. It is important to point out that the activity of neurons in animals, as opposed to electronic gates, is not binned in time and their response properties usually depend on their history. Both of these differences make the notion of a truth table as a sufficient representation of the neural responses with respect to the input space problematic in the case of a brain.

A currently running initiative, the human brain project, declared the goal to simulate the human brain. In case of its predecessor, the blue brain project, the simulation of a cortical column was performed [Markram, 2006]. These simulations were based on statistical models of connectivity based on single cell morphology and pairwise connectivity measurements averaged over animals and instances of neurons of a given cell-type [Markram et al., 2015]. These approaches for generating connectivity make implicit assumptions. Most notably the statistical models of connectivity



Nature Reviews | Neuroscience

Figure 5.2: **Function vs. structure** exemplified using the analogy of the brain circuits of neurons to a electrical circuits of logic gates. a) shows the truth table of the network of logic gates presented in b). Note that arrow inside the the zoom-in in a) highlights the only position where the truth table changes if the whole circuit marked with the dashed line in b) would have been omitted. Adapted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience [Denk et al., 2012], copyright 2012.



based on morphological reconstructions usually rely on some form of Peter’s rule [Peters and Feldman, 1976], which states that the probability of a connection between two neurites is proportional to some notion of spatial proximity. Limitations of this rule have been discussed [Binzegger et al., 2004] and evidence against this assumption has recently been presented [Kasthuri et al., 2015]. Furthermore it is important to point out the differences between mapping the complete set of connections within a brain region, sometimes called a dense reconstruction, and the sampling of pairwise connectivity between cell types. The sampling of pairwise connectivity is usually based on multi-patch recordings and averaged over animals and cell types [Markram et al., 1997a] [Feldmeyer et al., 1999] [Jiang et al., 2015]. In response to [Jiang et al., 2015] several authors [Barth et al., 2016] pointed out the general limitations of this approach like slicing artifacts and consistent definition of cell types. More importantly the approach of sampling pairwise connectivity cannot provide information about higher-order connectivity, which is defined as connections between pairs of cells whose probability depends on more than the morphological cell-type identity of the involved cells. An example could be a connection probability that depends on whether one of the involved cells has a specific connection (instead of morphology). This subdivision of morphologically defined cell types based on connectivity was recently discovered in the retina [Helmstaedter et al., 2013].

Therefore the self-imposed goal of the initiative to simulate the human brain based on the available data seems to involve invalidated assumptions and approximations in how connectivity was measured that have not yet been tested for their validity. This has generated some opposition in the neuroscience community [Helmstaedter, 2013] [Frégnac and Laurent, 2014]. Dense mapping of connectivity within a brain region and measurement of its variability over animals will provide important input to such initiatives in the future.

While research in recent decades has made available information about the structural, biophysical and genetic properties of single neurons, a full description of connectivity on the synaptic level has not yet been measured for any significant fraction of the brain in any mammalian species. Furthermore nobody has shown that a neural network based on statistical connectivity models based on single cell morphologies and pairwise connectivity can reproduce the behavior of the brain. Generating a full description of the connectivity within a region of the brain is the goal of the scientific field introduced in the next section.

## 5.3 Connectomics

The term "connectomics" was first introduced in 2005 as "comprehensive structural description of the network of elements and connections forming the human brain" [Sporns et al., 2005] and has since been used for diverse scales, methods and biological goals. The scales on which these connections are mapped vary from the macroscale using functional magnetic resonance imaging (fMRI) and diffusion tensor imaging (dTI) over the mesoscale using anterograde or retrograde tract tracing to the microscale using electron microscopy. Measurements on the macroscale are either measuring the correlation in the blood oxygen level dependent (BOLD) signal between brain regions or the macroscopic diffusion properties of water molecules in the brain. According to the definition above this would mean the "network elements" are usually brain regions and the "connections" are either correlation of BOLD signal between these brain regions or faster diffusion of water molecules between them. The former implies co-activation of the brain regions while the latter would suggest neuronal tracts, collections of neuronal fibers, between them.

The other two scales both measure connectivity based on neurons as "network elements". The main difference is that with the microscale methods one can map all chemical synapses made within a given imaged region while the light microscopic approaches have so far been unable to provide the necessary resolution and labeling method to reconstruct all neurons in a brain region and usually work with very sparse labeling approaches [Helmstaedter et al., 2008a] and thus only reconstruct a few neurons per animal. This currently limits the discovery of higher order connectivity to the electron microscopy (EM)-based microscale methods. Note that expansion microscopy [Chen et al., 2015], which shifts the problem of resolution from the microscope to the sample by enlarging the latter, in combination with super-resolution light microscopy [Hell, 2007] [Rust et al., 2006] might require changes to this statement at some point. This thesis will focus on the microscale cellular resolution [Helmstaedter, 2013] electron-microscopy aimed at dense reconstruction of brain regions.

First results in the field of EM-based cellular resolution connectomics were generated by Sydney Brenner and colleagues who completed the mapping of the nervous system of the nematode *C.elegans* in 1986 [White et al., 1986] and results about parts of the nervous system a decade earlier [Ward et al., 1975]. These efforts were substantial and based on annotating each electron micrograph manually and then assigning a label to each of these cross-section annotations consistently through the image series. A

complete reconstruction based partially on these initial results but for example fixing a major gap in the connectivity of ventral cord neurons has since been published [Varshney et al., 2011].

More recently electron microscopy based mapping of neuronal circuits has been used to map parts of the brain of larger animals. Note that most of the studies presented in this section have substantial limitations, which will be discussed in section 7.4.

Studies in the mouse retina have revealed dendrite specific wiring asymmetry from starburst amacrine cells (SAC) dendrites onto ganglion cells [Briggman et al., 2011], see Figure 5.3. This wiring asymmetry enhances direction selectivity in ganglion cells based on specific selectivity for centrifugal motion in SAC dendrites measured earlier [Euler et al., 2002] which in turn preferentially make output synapses onto direction selective ganglion cells whose null direction aligns with the centrifugal motion selectivity of a given dendrite.

Furthermore dense mapping of synaptic circuits in the inner plexiform layer of the retina discovered new morphologically defined bipolar cell types [Helmstaedter et al., 2013] and has shown that morphologically defined cell types can be further subdivided based on their synaptic connectivity [Helmstaedter et al., 2013]. The discovery of new morphological cell types is a result of dense mapping which could also have been achieved by sparse methods like light microscopy, but is made more convenient due to the complete mapping of all neurons in a given region. The subdivision on morphological cell types based on synaptic connectivity on the other hand would not have been possible using sparse methods. Further analysis performed on datasets from these initial studies [Kim et al., 2014], together with electrophysiological measurements [Baden et al., 2013] have also suggested a mechanism responsible for generating SAC dendrite direction selectivity.

Another example in mouse are the initial results concerning circuits in primary visual cortex. It has been shown that inhibitory neurons receive local excitatory input from cells with a diverse set of orientation selectivity [Bock et al., 2011]. Another study confirmed that pyramidal neurons with similar orientation selectivity preferentially form synapses with each other [Lee et al., 2016] confirming results obtained previously using light microscopy only [Ko et al., 2011].

On the other hand drosophila has been used as another main model organism in connectomics. The optical lobe as well as the mushroom body [Takemura et al., 2017] have been mapped extensively. In the optical lobe the first study of a single medulla column suggested a circuit for motion detection [Takemura et al., 2013] while the second

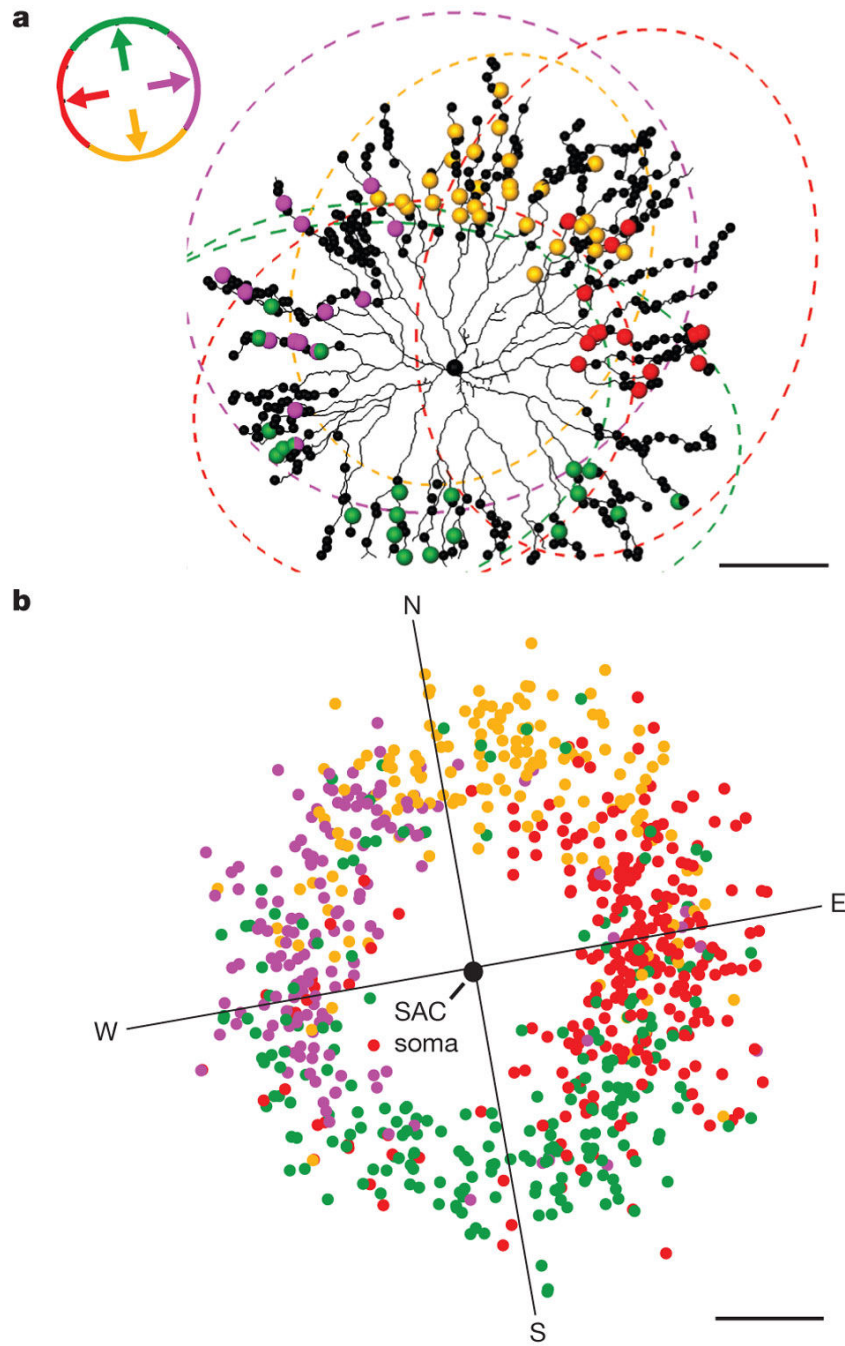


Figure 5.3: **Starburt amacrine cell output synapses** color coded according to direction selectivity of the postsynaptic direction selective ganglion cell. Scale bars: 50  $\mu\text{m}$  Reprinted by permission from Macmillan Publishers Ltd: Nature [Briggman et al., 2011], copyright 2011.

study found that wiring in the medulla columns of the fly visual system is very consistent between columns [Takemura et al., 2015]. The circuit that had been suggested for direction selectivity consisted of a very small spatial offset of Mi1 and Tm3 cells synapses onto T4 cells [Takemura et al., 2013]. In combination with information about the temporal delay between the responses in Mi1 and Tm3 interneurons [Behnia et al., 2014] in the medulla of the fly visual system, this suggested circuit has strong analogies to the generation of direction selectivity in the retina [Kim et al., 2014], where BC2 and BC3a types were observed to show differential time delay and make synapses with a spatial offset onto SAC dendrites. Note that recent data has shown that when the Tm3 cell is turned off only a subset of motion selectivity is disturbed and thus one might "require an elaboration of the currently prevailing model for ON motion detection" [Ammer et al., 2015]. See the section on "The Emergence of Direction Selectivity in T4-ON Cells" in [Mauss et al., 2017] for a detailed discussion of these limitations or [Borst and Helmstaedter, 2015] for a more general review about commonalities and differences of motion detection in the drosophila and mammal. For example the commonality of an early divergence of the ON and OFF responses that later synapse onto a single direction selective cell is discussed.

Another model system has been the fly larvae with initial studies showing higher path length density of dendritiform than varicose/globular neurites [Cardona et al., 2010], which is in stark opposition to the mammalian brain. Also the complete wiring diagram of the mushroom body of the drosophila larvae has recently been published [Eichler et al., 2017], which could lead to interesting comparisons to the adult circuits.

The olfactory bulb of the zebrafish larvae has also been reconstructed in full showing wiring specificity with respect to glomerular identity of the cell [Wanner et al., 2016]. Furthermore the zebrafish larvae is small enough to be feasible for whole brain electron microscopy today [Hildebrand et al., 2017].

Apart from these initial findings, method development is still required on the data acquisition as well as the data analysis to be able to map whole circuits in the cortex of mammals. The next section focuses on the data acquisition.

## 5.4 Electron Microscopy

As discussed in the last chapter currently the only viable method to obtain dense maps of higher order synaptic connectivity within a brain region is electron microscopy. Several methods for the acquisition of 3D image data using electron microscopy from

specifically prepared brain tissue have been developed over the past decades, see [Helmstaedter et al., 2008a] or [Briggman and Bock, 2012] for detailed reviews.

In order to obtain series of electron micrographs that can later be aligned into a 3D volume, a sample of brain tissue is acquired from the animal by either trans-cardial perfusion or immersion fixation after extraction. The advantages of the former approach are usually better preservation of the tissue while the latter procedure can be used in non-terminal experiments, for example on intra-operative human tissue. Afterwards an en-bloc staining procedure like reduced osmium thiocarbohydrazide-osmium (rOTO) [Willingham and Rutherford, 1984], sometimes used in conjunction with post-staining methods after cutting, are used to selectively increase the electron cross-section of bilipid membranes by introducing elements with a high atomic weight. For tissue blocks larger than 100 micrometer along their smallest dimension, improved versions of the en-bloc staining procedure were required and have been developed [Hua et al., 2015] [Mikula and Denk, 2015].

The two main types of electron microscopy used today are transmission electron microscopy (TEM) and scanning electron microscopy (SEM). TEM works on thin samples of tissue and the electrons traverse the sample. SEM is based on the detection of back-scattered or secondary electrons from the surface of the sample while scanning over the surface with an electron beam. The penetration depth and depth from which information about the structure will be collected when scanning over the surface in SEM varies with landing energy of the electrons and detector properties [Hennig and Denk, 2007].

The microscopy techniques can be further subdivided according to several properties of the cutting process used to generate imaging data from 3D volumes of tissue. One of the properties is whether the data is imaged before the respective piece of tissue is cut from the block, often termed block-face methods, or whether each slice is imaged separately after cutting.

The classic approach of serial section transmission electron microscopy (ssTEM), see Figure 5.4(a), falls into the latter category just like the approach of automated tape collecting ultra microtome (ATUM) [Hayworth et al., 2006] based SEM, see Figure 5.4(b). In both of these approaches the slices are cut prior to imaging, which has the advantage that the sectioning is independent of the imaging step, but introduces the challenge of transferring very thin slices onto some form of sample holder without damage or distortions. The classic approach of serial section transmission electron microscopy (ssTEM) has the main advantage of very high imaging speed

[Bock et al., 2011], especially when optimized for this purpose: "Net imaging throughput using the TEMCA2 system is 50 MPix/s" [Zheng et al., 2017]. The main improvement in the automated tape collecting ultra microtome (ATUM) based approach is that the collection of slices onto tape is automated, but as the carbon coated tape is not transparent to electrons, SEM has to be used.

The block-face imaging techniques image one side of the tissue block and afterwards remove a thin slice from the imaged surface of the block. The difference between diamond knife based serial blockface electron microscopy (SBEM) [Denk and Horstmann, 2004], see Figure 5.4(c), and focused ion beam based SBEM [Knott et al., 2008], see Figure 5.4(d), is the method used to remove the thin layer of tissue after imaging. In the former case the removal is achieved by automated cutting with a diamond knife while the latter approach uses ablation with a focused ion beam. Both approaches work on the block of tissue in an automated manner within the vacuum chamber of the electron microscope. Using an ion-beam has the advantage that higher doses can be used during the SEM based imaging of the sample, while the method is currently limited by the sample size that can be reliably ablated without image artifacts along the axis aligned with the ion beam. In general the block-face methods have the advantage that image deformations are less severe while the main challenge is that the cutting has to take place in the vacuum chamber of the microscope which can lead to contamination of the column or sample holder with debris from the cutting process which in turn can introduce image artifacts. An interesting recent development is the creation of a multi-beam scanning electron microscope (mSEM) by ZEISS [Eberle et al., 2015], which promises two order of magnitude higher imaging speeds than the currently used single beam scanning electron microscopes.

The work presented in this thesis is focused on a data set from layer 4 of primary somatosensory cortex obtained by trans-cardial perfusion from a mouse, followed by rOTO based staining as published in [Briggman et al., 2011] and diamond knife based SBEM acquisition by my colleague Kevin Boergens. The reasons for choosing this area are outlined in the next section.

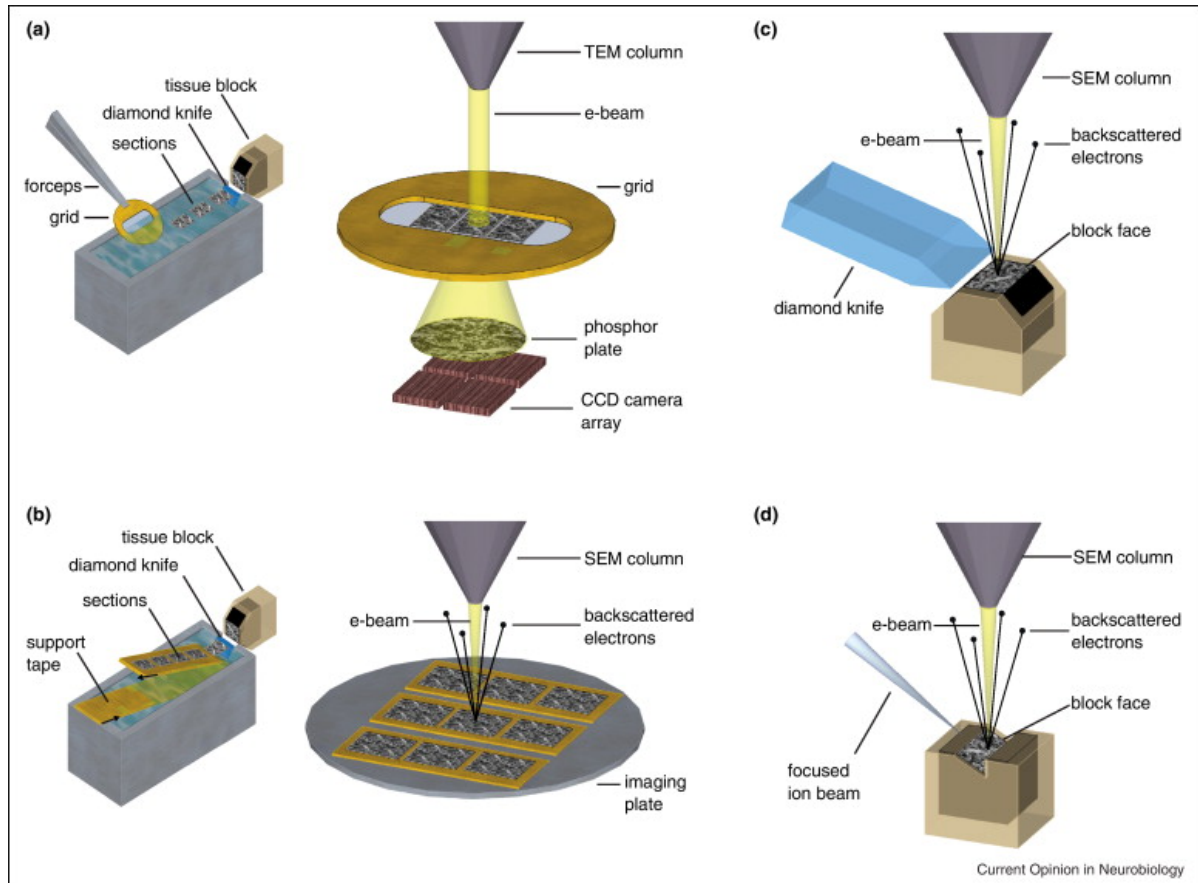
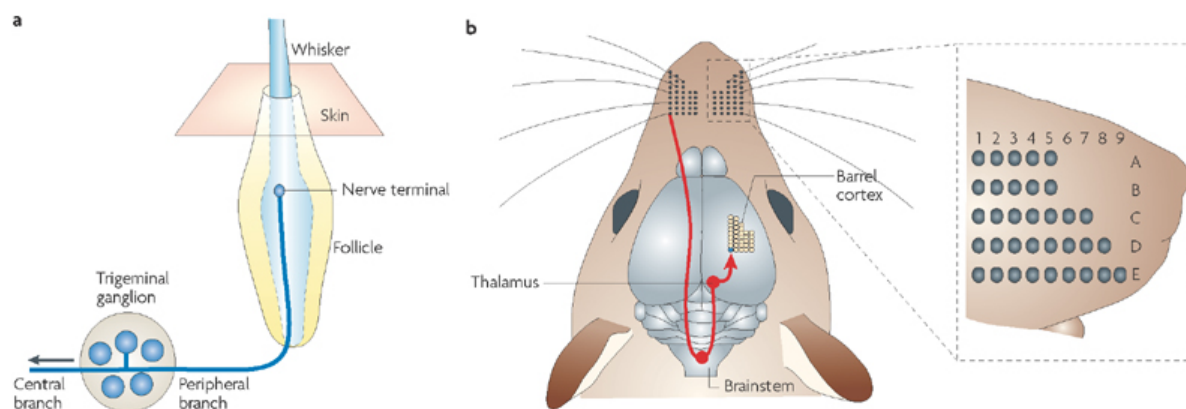


Figure 5.4: **3D electron microscopy approaches** (a) the classic approach of serial section transmission electron microscopy (ssTEM) (b) similar to (a) but with automated slice collection is automated tape collecting ultra microtome (ATUM) based scanning electron microscopy (SEM) (c) diamond knife cutting based en-bloc SEM (d) ion beam cutting based en-bloc SEM. Reprinted from [Briggman and Bock, 2012], Copyright 2012, with permission from Elsevier.



## 5.5 Barrel cortex

The brain region from which the main tissue sample presented in this thesis originated is called barrel cortex, which is a region with distinctive cytoarchitecture located in somatosensory cortex layer 4 of some rodent species. This region shows structures that have the 3D form of barrels when stained with e.g. the Nissl stain or for the presence of cytochrome c oxidase [Wong-Riley, 1989], with the bands in between these barrels called septa. The arrangement of these structures in cortex has the same layout as the whisker pad on the snout of the animal (see Figure 5.5). This structure was discovered in mice [Woolsey and Van der Loos, 1970] using the Nissl stain, while a more systematic study [Woolsey et al., 1975] found these structures in many other rodent species in the suborders Myomorpha (gerbils, hamster, rats, muskrat), Cavimorpha (chinchilla, guinea pig), Sciuromorpha (flying squirrel, chipmunk, ground squirrel, prairie dog, gray squirrel) and Hystricomorpha (african porcupine). It has since been shown that species from other orders also show such distinct structures, for example the star nosed mole [Catania and Kaas, 1997] and the wallaby [Waite et al., 1991]. The conservation of this structure in different species makes the research on this structure additionally interesting and will allow for cross-species comparison of local structural features.



Nature Reviews | Neuroscience

Figure 5.5: **The whisker sensory pathway.** The follicle of each whisker is innervated by the peripheral branch of a neuron in the trigeminal ganglion which in turn relays the signal via the thalamus to the barrel cortex. Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Neuroscience [Diamond et al., 2008], copyright 2008.

Figure 5.5 sketches the whisker sensory pathway that gives rise to these structures in cortex. The cell body of the first neuron in this pathway is located in the trigeminal

ganglion and one branch of this neuron innervates the follicle surrounding a whisker together with approximately 150 other neurons of the trigeminal ganglion [Vincent, 1913] and converts mechanical movement into an electrical signal. This signal will then travel past the cell body and down the other branch to relay the information to the trigeminal nuclei in the brainstem [Dörfl, 1985]. From there the information is relayed via either the nucleus ventralis posteromedialis (VPM) or posteromedial complex (POm) region in the thalamus to the primary and secondary somatosensory cortex. This means that only three synapses are activated, one in the trigeminal nuclei of the brainstem, one in the thalamus and one in the barrel cortex to relay the signal generated in the follicle of the whisker to cortical neurons. Note that in addition to the representation of the barrel field in cortex, the representation of the whiskers in the trigeminal nucleus as well as the representation in the VPM is preserving the topography of the whisker pad on the snout as well [Fox, 2008]. These regions are referred to as barrelettes and barreloids respectively in analogy to the barrel structures in layer 4 of primary somatosensory cortex.

The barrel cortex has been studied extensively using pairwise patch clamp recordings paired with morphological reconstruction, e.g. [Markram et al., 1997a] [Feldmeyer et al., 1999] [Helmstaedter et al., 2008b]. This has led to an accumulation of knowledge about the electrophysiological properties, single cell morphologies [Narayanan et al., 2015] as well as pairwise connection statistics between single cell types averaged over animals [Markram et al., 1997a] [Feldmeyer et al., 1999] [Helmstaedter et al., 2007]. Of particular interest when studying the barrel field in layer 4 is the connectivity between cells in layer 4, which was found to "suggest that in L4 of the barrel cortex synaptic transmission between spiny neurones is largely restricted to a single barrel" [Feldmeyer et al., 1999]. Another finding of particular interest when regarding the barrel cortex from a connectomics perspective is the description of an interneuron whose neurites are to a large extent restricted to layer 4 of a single barrel [Koelbl et al., 2013]. As long as connectomics is still limited by the size of the volumes that can be imaged with high quality, these two findings present an opportunity to reconstruct the complete connectivity between some cells when aiming to reconstruct a whole barrel.

Additional studies based on viral tracing and immunostaining have mapped the projections from thalamus to a barrel column [Wimmer et al., 2010], see Figure 5.6, the distribution of excitatory neurons in a barrel column [Meyer et al., 2010b], the distribution of boutons originating from nucleus ventralis posteromedialis (VPM)

and posteromedial complex (POm) within a barrel column [Meyer et al., 2010a] and the distribution of excitatory versus inhibitory neurons in the barrel column [Meyer et al., 2011].

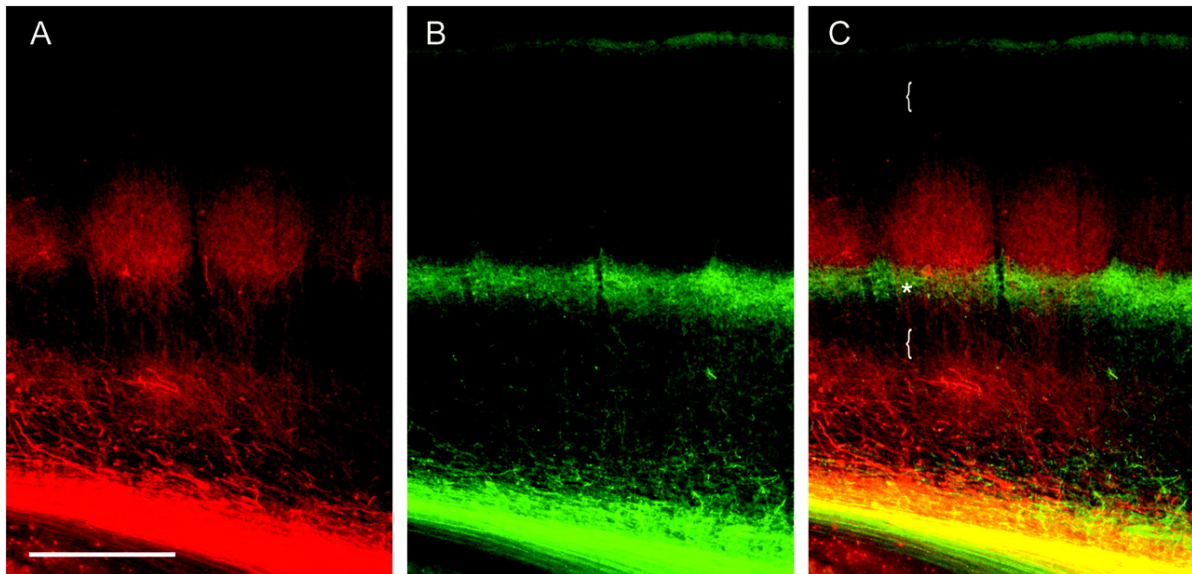


Figure 5.6: **Thalamocortical projections to the barrel cortex.** (A) Projections from the nucleus ventralis posteromedialis (VPM) to barrel cortex (B) Projections from the posteromedial complex (POm) to barrel cortex (C) Overlay of the both channels showing largely non-overlapping projections of the two thalamic nuclei in barrel cortex layer 4 and 5. Note that it also shows the columnar structure. This figure is reproduced from [Wimmer et al., 2010] and is licensed under the CC BY-NC 2.5.

The behavioral importance of whiskers has been proven for a long time, often in either elevated mazes, texture discrimination tasks [Vincent, 1912] or gap-crossing, which has been shown to be learnable by mice with only a single whisker [Celikel and Sakmann, 2007]. This finding is of particular interest, as combined with the structure of the whisker pathway and barrel field described above, one can hypothesize that a single barrel has behavioral relevance and can probably be altered in a learning paradigm. Similarly measurements in vivo of L2/3 cells in barrel cortex showed stronger and shorter onset latencies of cells in the barrel than in cells in the septum [Brecht et al., 2003]. More recent experiments have shown that whisking related activity in barrel cortex is higher during social facial touch compared to touching inanimate objects [Bobrov et al., 2014] and that active whisking strategies might change during different behavior [Arkley et al., 2014]. These results suggest that the simple stimuli and behavioral tasks used in many experiments might be a subsampling of the whole stimulus

space and might not represent the natural stimulus environment of these animals. In the analogy to a input-output table of a computational device introduced earlier one can again see that one of the challenges is finding a complete representation of the stimulus input space relevant in a certain brain region.

The extensive knowledge about the input pathway, single neuron properties, especially the confinement of the neurites of some cells to a barrel, behavioral relevance of a single barrel and its presence in multiple species and orders taken together make this a very interesting brain region for circuit mapping.

## 5.6 The reconstruction challenge

The development of better electron microscopy based 3D image data acquisition techniques described in an earlier section is rapidly advancing our ability to acquire 3D electron microscopy image data from larger volumes. Acquiring a dataset of layer 4 in barrel cortex from mouse, which has a rough size of  $(300\mu m)^3$ , is possible within approximately 1000 hours of experiment [Helmstaedter, 2013].

Using the conventional approach used by Sydney Brenner and colleagues of contouring all neurites for the dense reconstruction of such a stack would take an estimated 1 million human annotation hours [Helmstaedter, 2013]. These 3 orders of magnitude between the time needed for experiments vs. the manual data analysis workload has been termed the analysis gap [Helmstaedter, 2013].

Note that the example above only describes the reconstruction of one very small region of the mammalian brain. The work that would currently be involved in mapping the whole human brain with its "86.1  $\pm$  8.1 billion NeuN-positive cells" [Azevedo et al., 2009] or even the whole mouse brain with its "70.89  $\pm$  10.41 million total neurons" [Herculano-Houzel et al., 2006] is more understandable when contemplating that the barrel only contains  $4447 \pm 439$  neurons in rats [Meyer et al., 2010b]. When considering that this small set of neurons would already require 1 million annotation hours using the classic approach and that the effort would scale approximately linearly with the number of neurons, whole brain reconstruction seems far from feasible in mammals.

Note that the reconstruction challenge as described above focuses on cable reconstruction, but at least one other step is required for mapping the connectivity graph of a brain region: annotation of the connection between neurons, the synapses. As this is a local problem, a wrong classification of a synapse does not influence the classification

taken in another location, complete automation with a low error rate seemed more plausible in the near future than for wire reconstruction. [Staffler et al., 2017] describes an approach that makes it possible to perform automated synapse detection for a binary excitatory connectome with above 97% precision and recall.

The challenge of extracting the high-dimensional graph of neuron to neuron synaptic connectivity from densely packed neuropil involving wire reconstruction and synapse detection has been termed the "reconstruction challenge". Two possible solutions discussed in [Helmstaedter, 2013] are crowd-sourcing and machine learning. The next section will focus on the latter approach, while the former approach will be discussed in the last section of this introduction.

## 5.7 Artificial Intelligence & Machine Learning

Understanding intelligence is one of the main goals when studying the brain. An alternative approach to understand intelligence is creating artificial intelligence. In 1950 Alan Turing proposed a test for determining whether artificial intelligence has been achieved, which he called the "imitation game" [Turing, 1950]. This test suggests to test for artificial intelligence by asking a set of interrogators to chat with a human and a machine and judge which one of them is human. If the interrogators fail to distinguish the machine from the human, the machine could be deemed intelligent. This might be extended to other tasks to give intelligence a broader meaning and not focus on natural language processing.

As this goal has not yet been reached, even in the narrower sense, the field of "artificial intelligence" is today usually referred to as "machine learning", which seems to better describe what current algorithms are capable of. The most common form of machine learning is to learn one very specific task like detecting which of a given set of objects is present in an image based on a large amount of labeled examples [Krizhevsky et al., 2012]. This would be called supervised learning of image classification.

One of the possible solutions of the reconstruction challenge in EM based connectomics is to use methods from machine learning to automate (part of) the analysis workload. Interestingly people have also argued the other way around: that insights from neuroscience are required to further improve machine learning [Hassabis et al., 2017]. This is also the concept of the Intelligence Advanced Research Projects Activity (IARPA) Machine Intelligence from Cortical Networks (MICRoNS) program which aims to

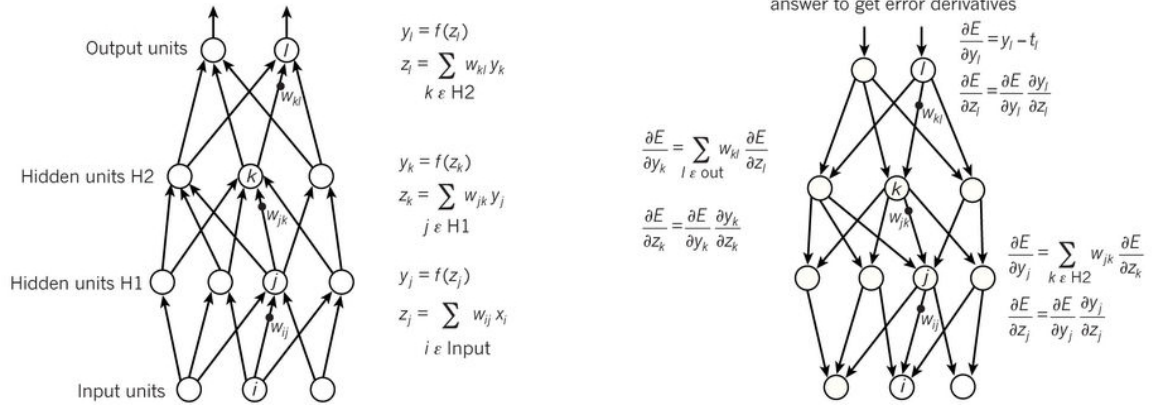
improve machine intelligence by studying cortical circuits and is currently a large funder of connectomics research. This parallel development within neuroscience and machine learning can be seen in the development of the field of connectomics in neuroscience to study connectivity in the brain and the field of connectionism in machine learning to study methods for fitting models to a certain task based on learning connectivity between simple units. This latter field is today often referred to as deep learning.

A first important step in the development of this field was to introduce the notion of a simple unit so that the connectivity between instances of these could be learned. This simple unit was first introduced as an element with multiple inputs and one binary output that activates if a given threshold on the summed inputs was surpassed [McCulloch and Pitts, 1943]. This was based on the observed "all or none" character of activity of neurons in our nervous system and therefore the simple units are often called neurons. First artificial neural networks made up of such simple units were developed a long time ago, with one of the more famous early examples being the perceptron [Rosenblatt, 1958].

More complex models of neurons than these simple units would be required to reproduce all biophysical properties of neurons. Nonetheless current machine learning methods rely on this definition of a simple unit with the slight alteration that the step-function in the example above is often replaced with other nonlinearities like e.g. sigmoids or rectified linear functions. It is thus important to keep the differences between neuron models in machine learning and biophysical plausible intracellular dynamics as e.g. in multi-compartmental modeling in mind. The artificial neuronal network models used in machine learning miss some other essential features observed in the brain, like neuron morphology and dynamic aspects of synapses.

The probably most important method developed for supervised learning is called backpropagation and was developed in the 1960's [Kelley, 1960, Dreyfus, 1962] for optimizing flight paths. Extensions of this technique are used in neural network implementations to date and it is the central algorithm used for supervised learning. The algorithm requires a set of input data with an associated label, which denotes the desired output. The model is then initialized into a state by drawing a random number for each parameter. From this state the model is updated according to the derivatives of the loss function with respect to the parameter during each sample presentation, see Figure 5.7 and [LeCun et al., 2015] for details.

A noticeable improvement in performance of machine learning algorithms for image classification tasks was observed with the introduction of a model termed convolutional



neural network (CNN). These have two important and related properties: weight sharing and translation invariance of the output. The translation invariance is achieved by weight-sharing, which means that some weights in the respective layers of a neural network are defined to be equal by design. This can be seen in analogy to the tiling of cell types in the retina which all extract similar features from their respective field of view. A similar property was found in the simple cells in the cortex of cats [Hubel and Wiesel, 1959]. The idea of translation invariance in pattern recognition was first used in the Neocognitron [Fukushima, 1979], while a well known implementation using max pooling layers for upsampling and backpropagation as a learning rule was described much later [LeCun et al., 1989]. This latter network managed to perform handwritten digit recognition with sufficient accuracy that it could be used for automatic zip code recognition. Other examples have since shown that these models can achieve better performance in many tasks, for example in image classification [Krizhevsky et al., 2012] or speech recognition [Hinton et al., 2012], than earlier neural network architectures or heuristic algorithms.

CNNs have been introduced to connectomics research with some modifications to use them as an intermediate stage for image segmentation algorithms [Turaga et al., 2009] [Turaga et al., 2010] [Jain et al., 2010]. Image segmentation is the problem of grouping pixels based on similarities in the local image content. For connectomics the goal is usually to group together pixels that belong to the same cell. For this purpose methods have been developed to train a CNN on a new cost function [Turaga et al., 2010] that allows one to directly optimize the rand index [Rand, 1971]. The Rand Index is one of the voxel-wise metrics used for judging segmentation quality. Another one that was developed for connectomics is called the warping error [Jain et al., 2010]. A network that directly optimizes the Rand Error was used for the connectomic reconstruction in [Helmstaedter et al., 2013] on image data stained to specifically suppress intracellular contrast.

The CNN implementation in [Berning et al., 2015] for classification of voxel in EM data into membrane vs. non-membrane voxel extended the traditional CNN definition used for images to 3D input data. The backpropagation was implemented similar to the definition given above. One of the difference to all earlier published approaches is that this model was trained on the conventionally stained EM data that provides the benefit of direct synapse identification based on intracellular structure, which in turn also makes the classification more difficult. This paper also presents the whole workflow for generating volume models using sparse annotations and introduces merger and split



rates as a skeleton based alternative to voxel based error metrics.

## 5.8 User interaction

Another possible solution of the reconstruction challenge in connectomics is using online crowd-sourcing to recruit large annotator crowds [Helmstaedter, 2013]. The existing in-browser annotation tool CATMAID [Schneider-Mizell et al., 2016] streams EM data by transmitting 2D chunks (images) of data. In [Boergens et al., 2017] a new approach of data delivery based on 3D cubes of data was implemented in a tool called webKnossos (<https://webknossos.org>). This data delivery is better suited to use the locality of data needed to annotate a certain region. It allows for arbitrary reslicing on the fly allowing for novel data interaction modes like "flight-mode", an egocentric single view-port data representation that allows faster tracing in a continuous movement paradigm. Rigorous quantification was provided that showed that error rates were indistinguishable between the conventional "ortho-" and the novel "flight-mode". webKnossos is based on the standalone application KNOSSOS [Helmstaedter et al., 2011] (<https://knossostool.org>).

Reconstructions based on skeletonization progressed at a speed of 3.7h/mm [Berning et al., 2015], 5.3h/mm [Helmstaedter et al., 2011], 7.69 [Helmstaedter, 2013] and 8.47h/mm [Wanner et al., 2016]. The new approaches presented in [Boergens et al., 2017] show that reconstruction with workload of less than 1h/mm is possible over a wide range of bandwidth and latency regimes of the internet connection the user has at his or her disposal.

# 6 Results

## 6.1 SegEM: Efficient image analysis for high-resolution connectomics

**Abstract** Progress in electron microscopy-based high-resolution connectomics is limited by data analysis throughput. Here, we present SegEM, a toolset for efficient semi-automated analysis of large-scale fully stained 3D-EM datasets for the reconstruction of neuronal circuits. By combining skeleton reconstructions of neurons with automated volume segmentations, SegEM allows the reconstruction of neuronal circuits at a work hour consumption rate of about 100-fold less than manual analysis and about 10-fold less than existing segmentation tools. SegEM provides a robust classifier selection procedure for finding the best automated image classifier for different types of nerve tissue. We applied these methods to a volume of  $44 \times 60 \times 141 \mu m^3$  SBEM data from mouse retina and a volume of  $93 \times 60 \times 93 \mu m^3$  from mouse cortex, and performed exemplary synaptic circuit reconstruction. SegEM resolves the tradeoff between synapse detection and semi-automated reconstruction performance in high-resolution connectomics and makes efficient circuit reconstruction in fully-stained EM datasets a ready-to-use technique for neuroscience.

**Contributions** All work was performed by M.B. and M.H.; K.M.B. acquired the cortex dataset.

**Copyright** Reprinted from [Berning et al., 2015], Copyright 2015, with permission from Elsevier.

# SegEM: Efficient Image Analysis for High-Resolution Connectomics

## Highlights

- SegEM: semi-automated reconstruction workflow for connectomics image data
- Reconstruction speed about 10-times higher than that of alternative methods
- Applied to 3D electron microscopy data from mouse retina and cortex
- Ready-to-use analysis code for novel datasets

## Authors

Manuel Berning, Kevin M. Boergens, Moritz Helmstaedter

## Correspondence

manuel.berning@brain.mpg.de (M.B.),  
mh@brain.mpg.de (M.H.)

## In Brief

Berning et al. developed SegEM, an image analysis workflow for efficiently extracting synaptic connectivity maps (connectomes) from large 3D-EM image datasets. The method is about 10-times more efficient than alternative methods.



# SegEM: Efficient Image Analysis for High-Resolution Connectomics

Manuel Berning,<sup>1,\*</sup> Kevin M. Boergens,<sup>1</sup> and Moritz Helmstaedter<sup>1,\*</sup>

<sup>1</sup>Department of Connectomics, Max Planck Institute for Brain Research, Max-von-Laue-Strasse 4, 60438 Frankfurt, Germany

\*Correspondence: [manuel.berning@brain.mpg.de](mailto:manuel.berning@brain.mpg.de) (M.B.), [mh@brain.mpg.de](mailto:mh@brain.mpg.de) (M.H.)

<http://dx.doi.org/10.1016/j.neuron.2015.09.003>

## SUMMARY

Progress in electron microscopy-based high-resolution connectomics is limited by data analysis throughput. Here, we present SegEM, a toolset for efficient semi-automated analysis of large-scale fully stained 3D-EM datasets for the reconstruction of neuronal circuits. By combining skeleton reconstructions of neurons with automated volume segmentations, SegEM allows the reconstruction of neuronal circuits at a work hour consumption rate of about 100-fold less than manual analysis and about 10-fold less than existing segmentation tools. SegEM provides a robust classifier selection procedure for finding the best automated image classifier for different types of nerve tissue. We applied these methods to a volume of  $44 \times 60 \times 141 \mu\text{m}^3$  SBEM data from mouse retina and a volume of  $93 \times 60 \times 93 \mu\text{m}^3$  from mouse cortex, and performed exemplary synaptic circuit reconstruction. SegEM resolves the tradeoff between synapse detection and semi-automated reconstruction performance in high-resolution connectomics and makes efficient circuit reconstruction in fully-stained EM datasets a ready-to-use technique for neuroscience.

## INTRODUCTION

Mapping neuronal circuits at single-cell resolution is the goal of high-resolution connectomics (Helmstaedter, 2013; Denk et al., 2012; Lichtman and Denk, 2011). While 3D-EM imaging methods have progressed substantially (Denk and Horstmann, 2004; Knott et al., 2008; Hayworth et al., 2006; Helmstaedter et al., 2013; Takemura et al., 2013; for a review, see Briggman and Bock, 2012), and 3D-EM imaging setups are being installed in many laboratories worldwide, the reconstruction speed of such data is lagging behind by at least 3 orders of magnitude (Helmstaedter, 2013). Fully automated reconstruction approaches have not provided the required reconstruction accuracy to date (Funke et al., 2012; Andres et al., 2012a, 2012b; Vazquez-Reina et al., 2011; Seyedhosseini et al., 2011, 2013; Jain et al., 2007, 2010a, 2011; Turaga et al., 2009, 2010; Ciresan et al., 2012; Kaynig et al., 2015; Nunez-Iglesias et al., 2014; Sommer et al., 2011; Liu et al., 2014). Instead, combinations of massive

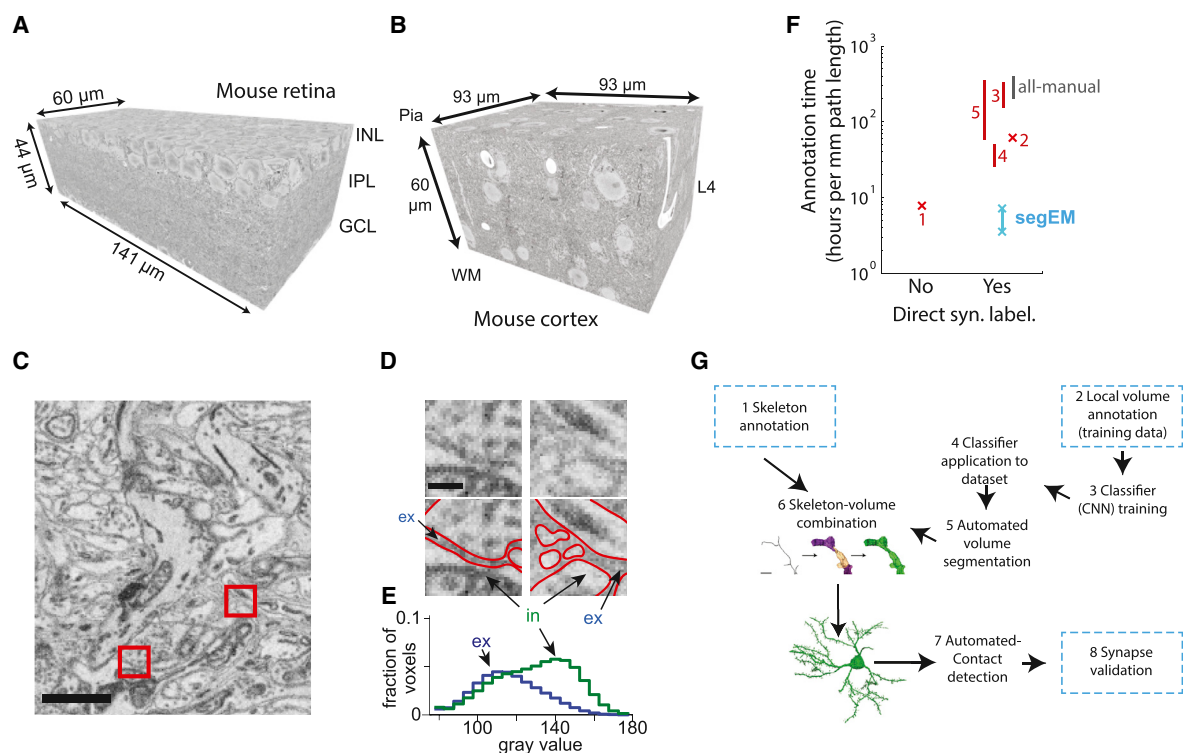
manual annotation with automated analysis methods have yielded first substantial connectivity maps in the fly optical system (Takemura et al., 2013) and mouse retina (Helmstaedter et al., 2013).

However, these approaches were limited either by analysis speed when proofreading pre-segmented data (Takemura et al., 2013; 14,400 hr investment for 105 mm circuit path length) or by the lack of direct synapse identification when using special cell-membrane-enhanced EM staining (Briggman et al., 2011; Helmstaedter, 2013; ~20,000 hr investment for 640 mm circuit path length).

In fact, even small blocks of neuronal tissue (Figures 1A and 1B) contain enormous amounts of local neuronal circuitry: for example, a block of mouse retina sized  $44 \mu\text{m} \times 60 \mu\text{m} \times 141 \mu\text{m}$  on a side contains about 400 cell bodies, 1.5 m of neuronal wires, and 0.5 million chemical synapses (Figure 1A; Briggman et al., 2011; Helmstaedter et al., 2013). To reconstruct neuronal circuits from such a sizeable volume of neuronal tissue, dendrites and axons of all neurons have to be followed through the dataset and all synapses identified. Unequivocal synapse identification requires the staining of synaptic vesicles and potential postsynaptic structures. Such conventionally stained 3D-EM data (Figures 1C and 1D), however, also stains mitochondria and other intracellular structures, which results in highly overlapping single-voxel gray value distributions (Figure 1E).

The fully manual volume reconstruction of neuronal circuits in such a 3D-EM dataset (Figures 1A and 1B) would consume enormous amounts of human work hours even for medium-sized circuits. The reconstruction of neuronal circuits between, for example, 100 input axons and 100 postsynaptic neurons in such a dataset from layer 4 of cerebral cortex (Figure 1B) would consume about 200,000–500,000 work hours, amounting to \$2–\$5 million resource investment, which makes such analysis prohibitive in most settings. It is therefore essential to develop automated classifiers for large-scale 3D-EM data that have been stained for all relevant structures, including synaptic vesicles and post-synaptic densities, and to integrate such automated classifiers into a reconstruction workflow that provides full-neuron volume reconstructions at a tolerable investment of manual labor.

To allow the analysis of large-scale EM data, several software tools have been developed that either focus on the fully manual annotation of neurites and synapses (KNOSSOS, Helmstaedter et al., 2011; TrakEM2, Cardona et al., 2012; CATMAID, Saalfeld et al., 2009) or provide a combination of automated analysis and proof-reading capabilities (rhoANA, Kaynig et al., 2015; ilastik, Sommer et al., 2011). While the fully manual tools cannot relieve



**Figure 1. SegEM-Based Connectomic Analysis of Fully Stained 3D-EM Datasets**

(A) Dataset boundaries of a  $44 \mu\text{m} \times 60 \mu\text{m} \times 141 \mu\text{m}$  sized stack acquired by SBEM in mouse retina (dataset ek563; Briggman et al., 2011) extending from the ganglion cell layer (GCL) via the inner plexiform layer (IPL) to parts of the inner nuclear layer (INL).  
 (B)  $93 \mu\text{m} \times 60 \mu\text{m} \times 93 \mu\text{m}$  sized SBEM dataset from mouse somatosensory cortex layer 4 (dataset 2012-09-28\_ex145\_07x2; K.M.B. and M.H., unpublished data). Orientation with respect to pia surface (Pia) and white matter (WM) is indicated.  
 (C) Neuropil was “fully stained” contrasting all membranes (retina).  
 (D) Magnified excerpts from (C) illustrate difficulty of distinguishing plasma membranes from vesicular and mitochondrial membranes. Outlines of plasma membrane boundaries from human annotation separating intracellular (in) and extracellular (ex) space are indicated in red. Excerpts sized  $(35 \text{ vx})^2$  each.  
 (E) Gray value distributions of extra- (blue) and intracellular (green) image voxels are highly overlapping, making automated analysis difficult.  
 (F) Annotation time estimates per mm neurite path length as a measure of circuit size (see also Figure 6A) shown for all-manual volume annotation, finished large-scale connectomic reconstructions (1: Helmstaedter et al., 2013; 2: Takemura et al., 2013), methods descriptions (3: RhoANA, Kaynig et al., 2015; 4: RhoANA applied to cerebral cortex data, Kasthuri et al., 2015; 5: Jones et al., 2015), and SegEM (reconstruction throughput for retina [top marker] and cortex [bottom marker]). See Supplemental Experimental Procedures for details of calculation.  
 (G) SegEM flowchart for the reconstruction of neuronal circuits. Blue dashed boxes: steps involving manual annotation. Note that neurite reconstruction (step 1) is by far the most time-consuming step (see also Figure 6A).

the enormous work load of reconstructing even small circuits, the computer-supported tools require semi-automated manual annotation or proof reading. The required residual proof-reading efforts are substantial, amounting to about 30–300 work hours per mm path length (for a single annotator, Figure 1F). This is up to 10 times faster than fully manual annotation, but still an enormous burden for circuit reconstruction for most laboratories.

If automated classifiers for fully stained 3D-EM data were available that could be directly combined with efficient skeleton reconstruction, an additional reconstruction efficiency gain of about 10-fold would be possible, which can propel the field into a realistic regime of reconstructing local neuronal circuits sized on the scale of 0.1–1 m integrated path length (Figures 1F and 6A).

Here, we report the development of such automated machine-learning-based classifiers and segmentation procedures that

operate on fully stained synapse-labeled large-scale volume EM data. We used convolutional neural network (CNN) classifiers since they are known to perform well in settings with little prior algorithmic knowledge about the classification task and since they profit directly from increased training set size. We applied semi-automated classifier selection routines that enable classifier training with little prior knowledge in a setting targeted to large-scale EM data applications.

Our classifiers achieve a volume segmentation quality that readily provides full-volume reconstructions when combined with skeleton-annotated neuron reconstructions, at an effective resource consumption of 4–8 hr/mm path length per annotator. We exemplify efficient circuit reconstruction for the bipolar cell inputs to one amacrine cell in an EM dataset from mouse retina, and local circuit reconstructions of spiny stellate cells and innervating axons in a dataset from mouse somatosensory cortex

layer 4. From these example reconstructions, we can predict that even circuits between hundreds of neurons are reconstructable using SegEM at resource investments that are realistic in most laboratories (Figure 6A).

SegEM (Data S8; [segem.io](http://segem.io)) integrates all tools required to perform circuit analysis in these datasets and the toolset for efficiently analyzing novel 3D-EM datasets of neuronal circuits. In addition, SegEM comprises all training data and test metrics (Data S8, [segem.io](http://segem.io)) to establish a large-scale 3D segmentation challenge for machine learning in connectomics.

In the following, the SegEM workflow and its application to two large 3D-EM datasets from mouse retina and cortex is described. Then, circuit reconstruction examples are presented for these two datasets, followed by general analysis guidelines for the application of SegEM to 3D-EM datasets. Finally, segmentation error rates are reported and discussed.

## RESULTS

We used SegEM to analyze two large 3D-EM datasets in which the neuronal tissue had been fully stained to resolve neurites and synapses. Both datasets were imaged using serial blockface scanning electron microscopy (SBEM; Denk and Horstmann, 2004): one from mouse retina IPL (ek563; Briggman et al., 2011; Helmstaedter et al., 2013; voxel size:  $12 \times 12 \times 25 \text{ nm}^3$ , dataset size:  $44 \times 60 \times 141 \mu\text{m}^3$ ; Figure 1A) and one from mouse S1 neocortex (2012-09-28\_ex145\_07x2, K.M.B. and M.H., unpublished data; voxel size:  $11.24 \times 11.24 \times 28 \text{ nm}^3$ , dataset size:  $93 \times 60 \times 93 \mu\text{m}^3$ ; Figure 1B).

### SegEM Workflow

The SegEM tools are aimed at replacing the labor-intensive manual volume segmentation of fully stained 3D-EM data by an efficient combination of skeleton annotation with automated volume segmentations (Figure 1F). The first branch of the workflow (Figure 1G) is the skeletonizing of axons and dendrites of interest using efficient 3D skeletonization software (such as KNOSSOS, Helmstaedter et al., 2011; [www.knossostool.org](http://www.knossostool.org), step 1 in Figure 1G). This is the key step that consumes manual labor. The consumption is about 4–8 hr per mm neurite path length for a single skeleton annotator, thus 25- to 100-fold faster than fully manual volume annotation (Helmstaedter et al., 2011; see Figure 1F).

To make use of this skeletonized data for volume segmentation, contact detection, and further circuit analysis, the second branch of the SegEM workflow is required (steps 2–5 in Figure 1G). First, small example volumes need to be manually labeled for training the automated classifiers (step 2 in Figure 1G). This step consumes about 1,000–2,000 work hours, but only once per dataset, which is only a fraction of the work hour investment required for neurite skeletonization in most settings. Next, the automated image classifiers have to be trained (step 3 in Figure 1G). Then, the trained classifier is applied to the 3D-EM image data, followed by an automated segmentation step (steps 4 and 5 in Figure 1G). The result of these SegEM steps is a piecewise volume segmentation, which is then combined with the skeleton reconstructions to yield full-volume reconstructions of neurons.

These full-neuron reconstructions are then fed into the SegEM-contact detection routines (step 7 in Figure 1G), which output contact area matrices and pointers to potential synapse locations. These can be used to finally validate synapses in the circuit of interest (step 8 in Figure 1G).

The SegEM workflow is rather modular: for example, the classifier training (steps 2 and 3 in Figure 1G) can be omitted if an existing classifier is to be applied to the data (for example, the classifiers developed here can be directly applied to novel 3D-EM datasets, see Data S2).

The logic of the SegEM workflow is described in the following as it was applied to the two large example datasets from retina (Figure 1A) and cortex (Figure 1B). A detailed step-by-step instruction for the application of SegEM can be found in the Supplemental Experimental Procedures.

### Training Data

Machine-learning-based image analysis requires substantial amounts of labeled training data. For the retina dataset, we used manually segmented image volumes in the surroundings of ribbon synapses, which consisted of bipolar, amacrine, and ganglion cell neurites (Figure S1A) as training data. These annotations had been used in a previous study for the calibration of synapse probability based on neurite-to-neurite contact area (Helmstaedter et al., 2013). The 215 manual segmentations were generated by contouring of neurites by 33 trained undergraduate students using a custom-written software (KLEE, implemented in MATLAB; other available tools are, for example, ilastik, Sommer et al., 2011; CATMAID, Saalfeld et al., 2009; VAST, Kasthuri et al., 2015) and consumed a total of about 1,500 hr annotation time (thus, about \$15,000 salary expense, which is on the scale of investments in chemicals for 3D-EM projects). This volume segmentation was split into a training and test set ( $\sim 5 \times 10^8$  and  $\sim 1 \times 10^6$  voxels, respectively).

For the cortex dataset, we volume-annotated locally dense data cubes of size  $(100 \text{ voxel})^3$ . We found training set size to be of critical importance. We initially trained with 10 of such volumes sampled evenly from the entire dataset (thus, about  $10^7$  training samples). However, classification results were much poorer than when training on the final 279 volumes ( $2.8 \times 10^8$  samples). Thus, the factor 30 in annotation cost and effort was likely required (total of 2,000 hr manual labeling time).

### Classifier Training

We trained convolutional neural network (CNN) classifiers to convert the raw 3D-EM image data into 3D maps of intracellular continuity between face-to-face adjacent voxels (one map for each cardinal direction; Figures S1A–S1C; Jain et al., 2010b; Turaga et al., 2010). For the cortex data, we found it sufficient to train the CNNs to output a single 3D map, thus representing the probability of each image voxel to be intra- or extracellular. Training the CNN to output one map is the default setting in SegEM but may be switched to three output maps for lower-resolution and lower-contrast image datasets.

To find optimal learning rates and architectural parameters for our CNN training, we implemented a simple hyperparameter search. This relieved us of the need to hand-design training parameters, and it made direct use of the availability of GPU



compute clusters without the need to explicitly parallelize the training computations. Instead, we used a GPU cluster for parallel training of different CNN architectures. This was essential for the success of the classifiers. While more sophisticated hyperparameter search algorithms have been proposed (Bergstra et al., 2013; Snoek et al., 2012), the simple evolutionary approach as described in the following was robust and successful and was aimed at making CNN training for connectomics widely applicable. When applying SegEM to a novel 3D-EM dataset, we recommend to first test our best-performing classifiers (Data S2) on the novel dataset. This is especially recommended if the high-resolution ultrastructure is comparable to cortex tissue, such as is the case for hippocampus, thalamus, and many other subcortical regions, and the imaging resolution is similar. Then, SegEM can be used to refine the classifier by training it using the best training parameters from our search (reported in Table 1). For novel datasets that have different ultrastructure and geometry, we recommend using the entire SegEM hyperparameter search procedure as described in the following.

The cost function for CNN classifier training was squared voxel-based error between the adjacency maps and the graded classifier output; the cost function was masked for unlabeled image regions. While this voxel-based metric is known to be conceptually inferior to segmentation-based metrics, we profited from its more efficient implementation yielding training times on the scale of days, not months (see below). Then we randomly selected training batches out of that masked volume. These training batches had to fulfil the criteria that at least 1/3 of voxels in the batch were labeled and that at least 1/3 of these labeled voxels were from each of the two label classes. These criteria were necessary to reduce training time and to avoid activity map saturation (such saturation could occur because the two label classes were highly unbalanced in regions containing large dendrites or somata).

For classifier training, two modifications of the training data proved crucial for classifier convergence and performance. (1) Since the detection of neurite borders is essential for avoiding incorrect neurite mergers, we reduced the bias toward intracellular regions by enlarging neurite-to-neurite walls in the labeled data. This was implemented by eroding the volume objects in the training data with a spherical structuring element of size  $3^3$  voxel, and thus a broadening of neurite-to-neurite walls by about 20–80 nm (Figures S1E and S1F). (2) We excluded those training volumes that yielded substantial classification performance decreases (examples in Figure S1J; 11 of 215 training volumes were excluded, Table S1; see Supplemental Experimental Procedures for a quantitative guideline on which training data to exclude). This is an unusual approach in machine learning, but proved critical for classifier convergence. The training volumes that were excluded mostly contained significant errors in the manual annotations—caused by insufficient annotator attention or artifacts in the image data. We recommend this training data co-optimization step in settings where intense curation of training data is not feasible because it would consume substantial additional annotation time.

The CNNs were implemented on a GPU compute cluster, which accelerated the parallel screening of CNN architectures. CNN architectures and training procedures comprised a total

of 17 varied parameters (most relevant: number of hidden layers, number of feature maps per hidden layer, filter size, learning rates for weights and biases and their respective learning rate decay, and batch size; see Table 1 and Figure S1D). Exhaustive parameter screening was not possible, since training convergence took at least about 2 days on one GPU card. We therefore first performed a qualitative manual network architecture selection. After an initial screening of about 200 network parameter sets, we started a semi-automated network selection procedure (Figures 2A–2C; shown for the retina dataset) with the best parameter set from the initial broad search (these parameters ranges are reported in Table 1; see Figure S2B for typical examples of rejected classifier outputs).

When we applied SegEM to the cortex dataset, we made a few improvements to the classifier search procedure, which we recommend for applying SegEM to novel datasets. First, we made use of a larger GPU cluster (28 GPUs), but CNNs were trained for only 1 day per iteration (Figure S2F). Then, the best half of the CNNs were selected based on averaged training error over the most recent 200 batch iterations, and the parameters for these were varied again (see Table 1 for best-performing network parameters and Supplemental Experimental Procedures for details). This procedure was iterated.

When applying the SegEM classifier search to a novel EM dataset, we recommend first trying to use the best-performing SegEM classifiers directly. When we applied the retina-trained classifier to cortex data, we obtained rather poor results (Figures S2D and S2E) and therefore applied the entire SegEM parameter search to the cortex dataset again. However, for datasets resembling cortex data or retina data in ultrastructure, the transfer of classifiers between datasets may be feasible.

After testing the existing classifiers on novel datasets, we recommend then performing SegEM hyperparameter search by initializing with our most successful hyperparameters (Table 1). This can provide a successful classifier within only about 10 such search iterations.

After classifier optimization, the best 2–5 output networks were used as input to the segmentation optimization (see below).

### Automated Volume Segmentation

The three adjacency maps obtained as classifier output for the retina dataset or the one output map obtained for the cortex dataset were then used in a watershed-based segmentation procedure to generate a space-filling volume segmentation (Figure 3). The segmentation procedure consisted of the following steps (Figure 3). (1) Morphological opening and closing by reconstruction using a spherical structuring element of radius  $r_{se}$ . This was intended to suppress mergers in the segmentation by removing small connecting bridges between otherwise unconnected image regions. This step provided improvements for the retina, but not the cortex data (see Table 2). (2) Generation of markers for watershed using a threshold or local minimum operation on the classifier output (parameters:  $\theta_{mg}$  and  $\theta_{hm}$ ), followed by connected components. In contrast to the retina classifier, for the cortex data, the  $h_{min}$  operation for marker generation performed better than thresholding. (3) Exclude markers smaller than  $\theta_{ms}$  voxels. (4) Marker-based watershed on the result of

**Table 1. Search Ranges and Best Parameters for SegEM-CNN Training**

|                                   | Retina Initial Screening (ranges)            |                 | Retina CNN Selection Iteration 1   |                         | Retina CNN Selection Iteration 2   |                         |                         | Retina CNN Selection Iteration 3 |   |                         | Cortex Automated CNN Selection           |                                |                                 |
|-----------------------------------|--|-----------------|------------------------------------|-------------------------|------------------------------------|-------------------------|-------------------------|----------------------------------|---|-------------------------|--|--------------------------------|---------------------------------|
|                                   | Min  | Max             | CNN1,j (ranges)                    | CNN1,1                  | CNN2,j (ranges)                    | CNN2,1                  | CNN2,2                  | CNN3,j (ranges)                  | CNN3,1                                    | CNN3,2 (best class)     | CNN (ranges)                             | 20130516T204040 <sub>8,3</sub> | 20131012T234219 <sub>41,3</sub> |
| $n_{HL}$                          | 1  | 8               | 4                                  | 4                       | 4                                  | 4                       | 4                       | 4                                | 4   | 4                       | 3–4                                      | 4                              | 4                               |
| $n_{fm}$                          | 5  | 30              | 10                                 | 10                      | 10                                 | 10                      | 10                      | 10                               | 10  | 10                      | [15,15,10,10], [10,10,10,10], [10 10 10] | [10,10,10,10]                  | [10,10,10,10]                   |
| $n_{om}$                          | 1  | 3               | 3                                  | 3                       | 3                                  | 3                       | 3                       | 3                                | 3   | 3                       | 1  | 1                              | 1                               |
| $S_{Fx/y}, S_{Fz}$                | 5,5–20,10 <sup>a</sup>                       |                 | 8,4                                | 8,4                     | 8,4                                | 8,4                     | 8,4                     | 8,4                              | 8,4                                       | 8,4                     | 5,3–21,11                                | 11,5                           | 11,5                            |
| $d_b$                             | 1  | 5               | 2                                  | 2                       | 2                                  | 2                       | 2                       | 2                                | 2   | 2                       | 4  | 10                             | 4+                              |
| $S_{Bxy,z}$                       | 1,1–100,50 <sup>b</sup>                      |                 | 12,6                               | 12,6                    | 12,6                               | 12,6                    | 12,6                    | 4,2–100,50 <sup>c</sup>          | 12,6                                      | 4,2                     | 100,100                                  | 100,100                        | 100,100                         |
| $\eta_{w0}$                       | 10 <sup>-8</sup>                             | 10 <sup>3</sup> | 10 <sup>-5</sup> –10 <sup>-1</sup> | 10 <sup>-4</sup>        | 10 <sup>-6</sup> –10 <sup>-3</sup> | 10 <sup>-6</sup>        | 10 <sup>-6</sup>        | 10 <sup>-7</sup> –1              | 10 <sup>-7</sup>                          | 10 <sup>-6</sup>        | 10 <sup>-7</sup> –10 <sup>-13</sup>      | –                              | –                               |
| $\eta_{b0}$                       | 10 <sup>-6</sup>                             | 100             | 10–100                             | 10                      | 0.01–10                            | 0.1                     | 10                      | 0.01–1                           | 0.1                                       | 0.1                     | 10 <sup>-7</sup> –10 <sup>-13</sup>      | –                              | –                               |
| $\tau_w, \tau_b$                  | 10 <sup>-7</sup> –9 × 10 <sup>12</sup> (l,e) |                 | 5 × 10 <sup>8</sup> (l,e)          | 5 × 10 <sup>8</sup> (e) | 5 × 10 <sup>8</sup> (l)            | 5 × 10 <sup>8</sup> (l) | 5 × 10 <sup>8</sup> (l) | 5 × 10 <sup>8</sup> (l)          | 10 <sup>-7</sup> –5 × 10 <sup>8</sup> (l) | 5 × 10 <sup>8</sup> (l) | 5 × 10 <sup>8</sup> (l)                  | –                              | –                               |
| $N_{it,cum}$ (× 10 <sup>5</sup> ) | 5 × 10 <sup>-4</sup>                         | 3.7665          | 1.4355–                            | 1.7475                  | 2.806–                             | 2.8075                  | 2.8085                  | 2.8295–                          | 3.2055                                    | 3.3855                  | 3.5–7 × 10 <sup>-4</sup> / it.           | –                              | –                               |
|                                   |  |                 | 1.75                               |                         | 2.8095                             |                         |                         | 3.3855                           |   |                         |  |                                |                                 |

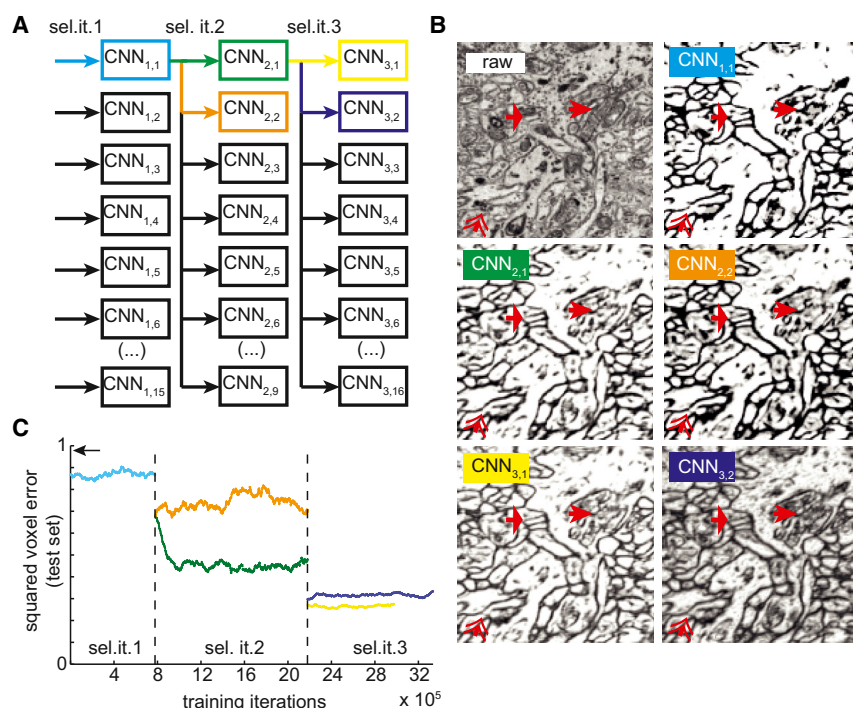
Network CNN<sub>3,2</sub> was best performing on retina data and used for Figures 3 and 4. Network 20130516T204040<sub>8,3</sub> was best performing on cortex data and used for Figures 3 and 5.  $n_{HL}$ : number of hidden layers;  $n_{fm}$ : number of feature maps per hidden layer;  $n_{om}$ : number of output maps;  $S_{Fx/y}, S_{Fz}$ : filter size in x/y and z direction in vx, respectively;  $d_b$ : mask border size in vx;  $S_{Bxy,z}$ : batch size in x/y and z direction, respectively;  $\eta_{w0}$ : weight learning rate initialization;  $\eta_{b0}$ : bias learning rate initialization;  $\tau_w, \tau_b$ : decay constants for weight and bias learning rates;  $N_{it,cum}$ : cumulative number of training iterations (number of batch learning iterations). l,e: linear and exponential learning rate decay, respectively.

<sup>a</sup>8,4; 20,10; 12,6; 10,5; 5,5; 7,7.

<sup>b</sup>[[5,5,5], [10,10,10], [2,2,2], [20,20,20], [1,1,1], [30,30,30], [12,12,6], [8,8,4], [4,4,2], [20,20,10], [40,40,20], [100,100,50]].

<sup>c</sup>[[12,12,6],[40 40 20],[20 20 10],[4 4 2],[100,100,50]].





**Figure 2. Classifier Training and Semi-automated CNN Hyperparameter Selection Procedure**

(A) Sketch of CNN variation and selection procedure (Sel.it: selection iteration). See Table 1 for CNN and training parameters.

(B) Example classification results (average affinity maps) for classifiers developed via procedures in (A). Arrows indicate typical classification challenges and their improvement over selection iterations (mitochondria close to plasma membrane, broad arrow; intracellular clustered large mitochondria, narrow arrow; intracellular staining precipitate close to plasma membrane and proximity to small diameter neurites, double arrow).

(C) Normalized test error development during selection procedure shown in (A). Colors match CNNs in (A) and (B). Error was calculated every 500 batch iterations and averaged over 101 of those errors using a sliding window. Arrow: initial error after random initialization of CNN.

segmentation step 1 (see Supplemental Experimental Procedures for details and Data S8 for all required routines).

Thus, the segmentation procedure had a total of three relevant parameters ( $[r_{se}, \theta_{mg}, \theta_{ms}]$  and  $[r_{se}, \theta_{hm}, \theta_{ms}]$ , respectively; Table 2). We optimized these segmentation parameters by an iterated parameter grid search (see Table 2 for parameter ranges and best-performing parameters). Segmentations were judged by the average inter-error distance (i.e., the combined split-merger inter-error distance, see below), the average segmentation object size, merger suppression, and avoidance of longitudinal neurite splits (see Experimental Procedures for details).

### Combining Skeletons and Automated Segmentations

We then used our automated volume segmentations for whole-neuron reconstruction (step 6 in Figure 1F). We skeletonized the axons of 276 bipolar cells and the dendritic tree of one amacrine cell from the mouse retina dataset ek563 (Figure 4). 37 reconstructions were identified to be type 5 cone bipolar cell axons based on their lamination in the inner plexiform layer (dendrites could not be reconstructed since ek563 only spanned from the GCL to the beginning of the INL, see Figure 1A; each axon was traced by one experienced student). Of these skeletons, 10 were randomly chosen (Figures 4A and 4B), proofread by one additional tracer, and all volume segmentation objects overlapping with at least one skeleton node were collected (inset in Figure 1F) for each bipolar cell axon. Skeletons were traced at an average consumption of  $7.2 \pm 4.1$  hr/mm, which is similar to the speed achieved in the surface-enhanced contrast sample from mouse retina (e2006; Helmstaedter et al., 2011, 2013; see Figure 1F for comparisons).

The resulting whole-cell volume reconstructions (Figures 4A and 4B) contained 3 obvious merge errors in 10 bipolar cells

with a total path length of 4.13 mm, i.e., 1.38 mm distance between mergers (see below for a more detailed quantification of error rates). This was comparable to

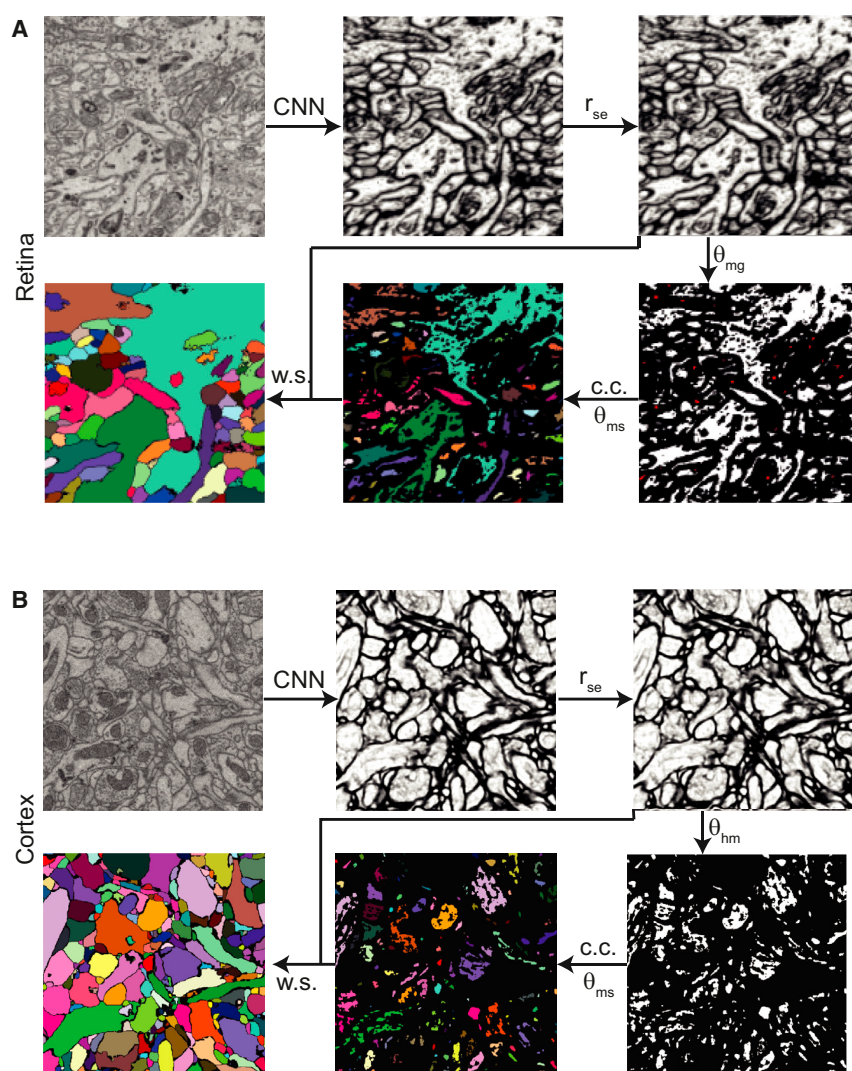
the error estimates in the retinal circuit reconstruction based on special cell-membrane enhanced staining.

### Automated Contact Detection and Synaptic Circuits in Mouse Retina

We then used the volume reconstructions of type 5, 6, and 7 bipolar cells (CBCs 5, 6, 7;  $n = 37, 37, 22$ ), rod bipolar cells (RBCs,  $n = 117$ ), yet unclassified cells ( $n = 23$ ), and a peculiar amacrine cell with wide-field ramifications (Figure 4C) to measure the synaptic bipolar-cell innervation profile of this amacrine cell. A total of 243 contacts between these neurons were automatically detected (step 6 in Figure 1F). Since we had previously shown that contact area predicts synaptic contact in bipolar-to-ganglion cell and amacrine cell synapses in mouse retina (Helmstaedter et al., 2013), we only validated three of these automatically detected contacts (of which two were ribbon synapses; Figure 4D). Synapse validation was performed using either KNOSSOS (Helmstaedter et al., 2011; knossostool.org) or web-KNOSSOS (K.M.B. M.B., T. Bocklisch, and M.H., unpublished data; webknossos.brain.mpg.de); SegEM outputs skeleton files that point directly to the contact locations of interest, thus facilitating synapse validation for the human annotator (step 7 in Figure 1F). Synapse validation consumed about 3–5 min inspection time per contact. Figures 4C and 4D show the resulting innervation pattern of the investigated amacrine cell, providing evidence for a spatially segregated CBC innervation (Figure 4D; total annotation time 280 hr: 1 hr for each bipolar cell axon, 3 hr for AC, and 3 min per inspected synapse).

### Synaptic Circuits in Cortical Layer 4

To perform local circuit analysis in cortex, we skeleton-reconstructed the dendrites of 4 spiny stellate neurons (52 hr of



**Figure 3. Segmentation Procedure Shown for Retina and Cortex Data**

(A) Raw data forward-passed through a CNN classifier (retina; classifier  $CNN_{3,2}$ , see Table 1) yielding three voxel-to-voxel affinity maps (only x-output  $[o_x]$  shown, see Figures S1A and S1C), followed by image inversion, morphological reconstruction with radial structuring element of radius  $r_{se}$  (shown is result for  $r_{se} = 1$ , see Table 2), map binarization at threshold  $\theta_{mg}$  (shown  $\theta_{mg} = 0.3$ ), application of connected component (c.c.) segmentation (26-surround) for marker generation, exclusion of markers with volume  $< \theta_{ms}$  voxels ( $\theta_{ms} = 150$  shown), and marker-based watershed (w.s.) on inverted morphologically reconstructed affinity maps using the shown markers.

(B) Classification and segmentation for cortex dataset; steps are as for retina data (A) except for local minimum operation with minimal depth parameter  $\theta_{hm} = 0.39$  ( $r_{se} = 0$ ,  $\theta_{ms} = 50$ ). Note that maps were inverted before watershed marker generation but are shown in non-inverted form here. Image sizes:  $(6.2 \mu m)^2$  (A) and  $(5.8 \mu m)^2$  (B).

per suspected synaptic contact. The resulting local high-resolution synaptic connectome (Figure 5D, bottom) thus required a total manual annotation effort of 53 hr; skeleton tracing was possible at only 3.7 hr per mm path length, which we attribute to the better data quality in our cortex dataset and the skill level of the annotators. We finally searched all pairs of axons and dendrites for additional synaptic contacts to estimate the rate of missed synapses by SegEM. We found no additional synaptic contact (0 missed contacts for 33 detected contacts), which provides an additional validation of the quality of the SegEM volume segmentations.

tracing), combined them with the SegEM-generated volume segmentations (Figures 5A and 5B), and then identified the axons providing synaptic input to three directly adjacent spines on one of the neurons' dendrites (cell 1, primary dendrite 1; Figure 5C). We then reconstructed the three corresponding pre-synaptic axons and used SegEM to compute the local high-resolution contact matrix (Figure 5D) reporting the size of all contacts between these three axons and all of the spiny neuron dendrites (total of 33 contacts). For neuron-to-neuron contacts in cortex, a direct prediction of the existence of a chemical synapse based on the size of an individual neuron-to-neuron contact is not possible. We therefore inspected each of the detected contacts and determined the local synaptic connectome (Figure 5D, bottom; 8 of 33 contacts were confirmed to be synaptic; examples in Figure 5E). Synapses were validated by visual inspection using direct links to contact locations in our web-based annotation tool, webKNOSSOS (K.M.B., M.B., T. Bocklisch, and M.H., unpublished data; [webknossos.brain.mpg.de](http://webknossos.brain.mpg.de)), which consumed less than 1 min annotation time

### Circuit Reconstruction Time Estimates

How realistic are synaptic circuit reconstructions at the measured throughput using SegEM, based on the small example circuits reported for the retina and cortex? Take, for example, the innervation of excitatory neurons in cortical layer 4 by thalamocortical axons (Figure 6A). While single synaptic innervations have been studied, the pattern of local target selectivity (or the lack thereof) of the thalamocortical innervation is not known. Therefore, a study to reconstruct the connectome between say 10 thalamocortical axons and 30 postsynaptic L4 neurons would be highly relevant. What investment would such a circuit reconstruction require using SegEM? Each of the involved neurites has a couple of millimeters of path length; let's assume that a redundancy of 4 tracings per dendrite and 2 per axon is about sufficient (tracing redundancy can be scaled to the required circuit accuracy, which depends

**Table 2. Best-Performing Segmentation Parameters for Retina and Cortex Segmentations**

|                                      | Segmentation Parameters Retina <sup>a</sup>           |   |                                     |                | Segmentation Parameters Cortex <sup>b</sup>           |   |                                     |                |
|--------------------------------------|---|---|-------------------------------------|----------------|---|---|-------------------------------------|----------------|
|                                      | Parameter ranges for split-merger metrics (Figure 7F) | Optimum segmentation (Figure 7F)<br>$\theta_n = 1; 2$ | Whole-cell segmentations (Figure 4) | All else shown | Parameter ranges for split-merger metrics (Figure 7F) | Optimum segmentation (Figure 7F)<br>$\theta_n = 1; 2$ | Whole-cell segmentations (Figure 5) | All else shown |
| $r_{se}$                             | {0,1}   | 1; 1  | 1                                   | 1              | 0   | 0; 0  | 0                                   | 0              |
| $\theta_{mg}$                        | {0.2,0.21,...0.5}                                     | 0.5; 0.38   | 0.31                                | 0.3            | –   | –   | –                                   | –              |
| $\theta_{ms}$                        | {0,20,40,60,100,150}                                  | 100; 0  | 20                                  | 150            | {0,50,100}  | 50; 0   | 10                                  | 50             |
| $\theta_{hm}$                        | –   | –   | –                                   | –              | {0.02,0.04,...0.7}                                    | 0.58; 0.04  | 0.25                                | 0.39           |
| IED ( $\mu\text{m}$ ) <sup>c</sup>   | –   | 3.85; 7.91  | 2.20; 7.04                          | 2.21; 6.24     | –   | 1.58; 4.93  | 0.93; 4.48                          | 1.26; 3.86     |
| Split ( $\mu\text{m}$ ) <sup>c</sup> | –   | 7.71; 9.76  | 2.37; 7.38                          | 2.56; 6.65     | –   | 2.89; 5.03  | 0.96; 4.56                          | 1.46; 4.47     |
| Merge ( $\mu\text{m}$ ) <sup>c</sup> | –   | 7.71; 41.73   | 30.26; 151.28                       | 16.36; 110.02  | –   | 3.48; 214.52  | 28.41; 241.52                       | 9.29; 28.41    |

$r_{se}$  = radius of spherical structuring element for morphological operations;  $\theta_{mg}$ : threshold used for marker generation;  $\theta_{hm}$ : depth parameter of H-Minima operation;  $\theta_{ms}$ : voxel threshold for watershed marker size. IED: average inter-error distance; split: average distance between splits; merge: average distance between mergers (see Experimental Procedures for calculation).

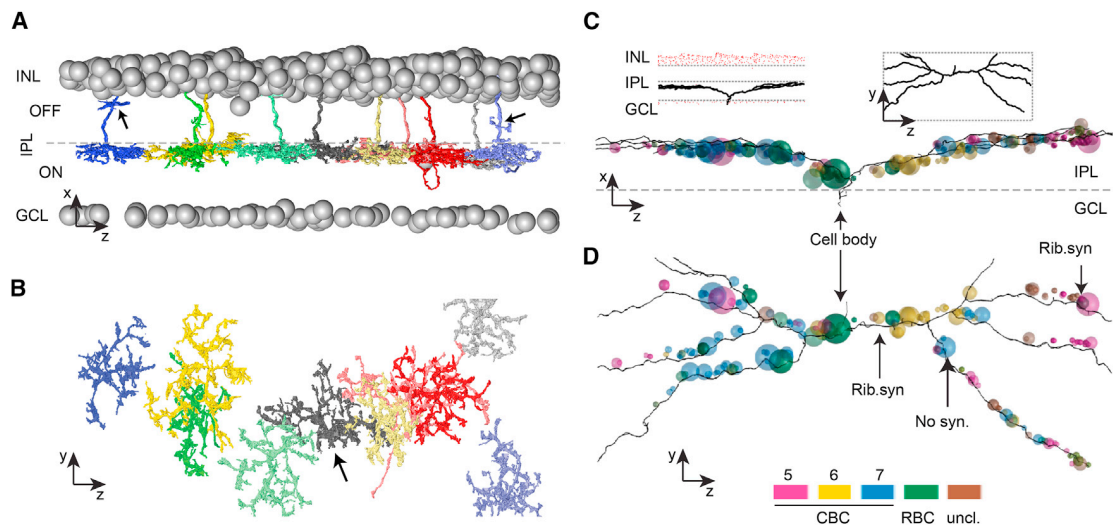
<sup>a</sup>Using CNN<sub>3,2</sub> (see Table 1).

<sup>b</sup>Using CNN 20130516T204040<sub>8,3</sub> (Table 1).

<sup>c</sup>Reported for node overlap thresholds  $\theta_n = \{1; 2\}$ , see Figure 7F.

on the type of circuit that is being studied; see Helmstaedter et al., 2011, 2013), then we end up with about 3,300 work hours for neurite tracing. If distributed between, for example, 15 undergraduate students, who each work about 40 hr per month, such an analysis would take about 5.5 months and consume

about \$33,000 (assuming about \$10 per hour salary expenses). Then, the about 1,000 contacts between the axons and dendrites need to be inspected, which would require about 15–20 work hours, thus only a fraction of the neurite reconstruction effort. Together, such a project is realistic in resource and work



**Figure 4. Connectomic Analysis in Fully Stained Retina Data**

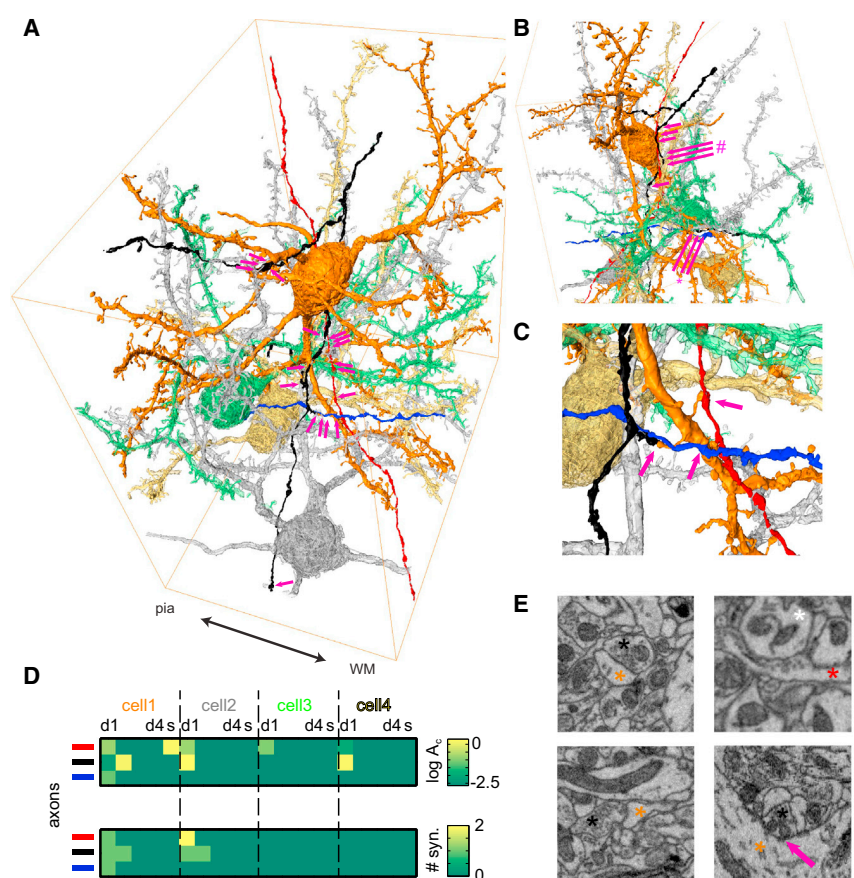
(A) Volume reconstruction of 10 randomly selected cone bipolar cell type 5 (CBC5) axons in dataset ek563 from mouse retina (see Figure 1) obtained by relabeling the computer-generated segmentation objects (classifier CNN<sub>3,2</sub>,  $r_{se} = 1$ ,  $\theta_{mg} = 0.31$ ,  $\theta_{ms} = 20$ , see Tables 1 and 2) according to human-generated skeleton annotations. Tracing time was about 1 hr per bipolar cell axon. Arrows point to obvious merge errors (3 obvious errors, total skeleton path length: 4.13 mm, i.e., inter-merger distance  $\sim 1.38$  mm). Gray spheres indicate somata.

(B) Same reconstruction as in (A) viewed in the IPL plane.

(C) Automated contact detection of CBC inputs to a displaced amacrine cell (insets in C). Contacts with CBCs of type 5–7, RBCs, and unclassified neurons are shown as spheres; sphere surface proportional to contact size. Since this dataset was fully stained, automated contacts can be visually inspected and existence and directionality of synapses confirmed (Rib. syn.: ribbon synapse CBC→AC confirmed; No syn.: accidental contact without evidence for synaptic contact). CBCs were annotated twice, AC once by an expert annotator. Annotation consumption: 7.2 hr/mm path length for a single annotator.

(D) Same reconstruction as in (C) viewed in the IPL plane. All images show the whole dataset with dimensions 141  $\mu\text{m}$  (z), 44  $\mu\text{m}$  (x), 60  $\mu\text{m}$  (y).





**Figure 5. Connectomic Analysis in Fully Stained Cortex Data**

(A) Volume reconstruction of 4 spiny neurons (orange, gray, green, yellow, respectively) by combination of skeleton tracing with SegEM classifier (see Figure 4, CNN 20130516T204040<sub>8,3</sub>; see Tables 1 and 2 for classification and segmentation parameters).

(B) Same reconstruction as in (A) viewed from a different angle.

(C) The presynaptic axons from three neighboring spines of one neuron (orange, C) were also volume reconstructed (black, blue, red axon). Arrows in (A), (B), and (C) point to all detected contacts between these axons and spiny neurons (\* and #: 5 and 6 contacts, respectively).

(D) The local connectome between the 3 axons (red, black, blue, respectively) and the primary dendrites (d1–d4) and somata (s) of the 4 spiny neurons. Top panel shows the total size of contact areas  $A_c$  (see magenta arrows in A–C). Bottom panel shows the result of visual inspection of each contact to determine the presence of chemical synapses.

(E) Examples of synapses and one accidental contact (bottom right). Annotation consumption: 3.7 hr/mm path length for a single annotator.

hour consumption and would provide unprecedented circuit data that currently cannot be obtained with other techniques. Notably, without SegEM, this reconstruction would have consumed about 10 times more work hours and resources, which would make it not feasible in most settings.

Similarly, one can project the required investments for other circuit reconstructions of similar or larger scale in cortex and retina (for example, the connectome between 300 bipolar and 300 small-field amacrine cells would consume about 6,500 hr total; see Supplemental Experimental Procedures and Figure 6A for details of the calculation and further examples).

#### Analysis Guideline for Novel Large-Volume EM Datasets

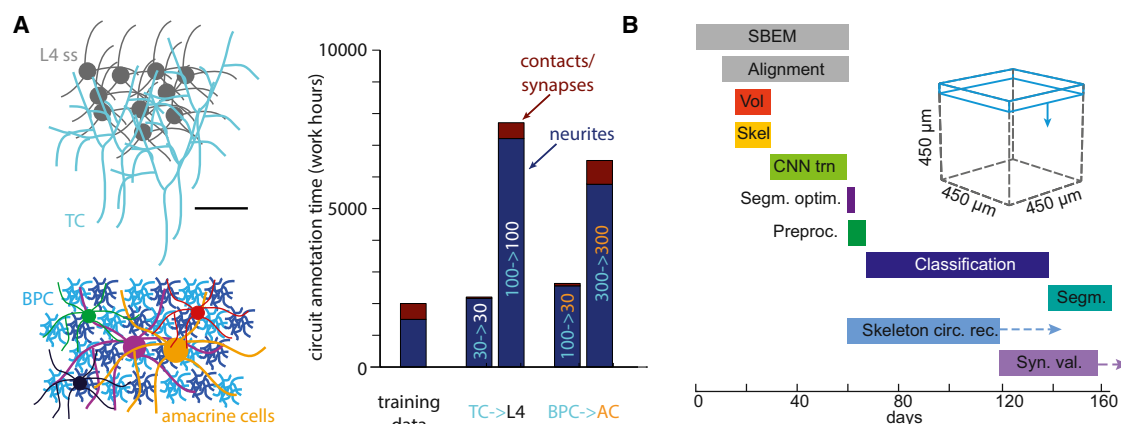
The analysis steps needed to perform circuit reconstruction in large-scale 3D-EM datasets (Figure 6B) can thus be summarized as a methodological guideline based on the SegEM analysis package. As a prototypical example, assume a SBEM dataset from mouse barrel cortex covering one layer 4 barrel from one cortical column at the required resolution (450  $\mu\text{m}^3$  volume,  $12 \times 12 \times 25 \text{ nm}^3$  voxel size, i.e., 23 TB image data, about 60 days of experiment for 16,000 successive image layers). Manual analysis for training label generation can be done in parallel to data acquisition (Figure 6B), such that the CNN selection procedures are finished when data acquisition is. Dataset classification and segmentation will then approximately equal the imaging time (assuming a compute cluster of  $\sim 60$  GPUs and

without additional parameter search (see above). Circuit reconstruction by skeleton reconstruction is then possible at a consumption of about 4–8 hr/mm path length. Within 2 months (as proposed in Figure 6B), a team of 30 annotators working at 40 hr/month can provide 2,400 work hours—sufficient for reconstructing the circuits between 30 TC axons and 30 spiny stellate neurons, or 100 bipolar cells and 30 amacrine cells (Figure 6A; 80–180 mm circuit size) at moderate cost. Of course, this analysis phase can be extended for larger circuits. All processing code required for such an analysis setup is contained in SegEM (segem.io, Data S8).

#### Segmentation Metrics

The main contribution of SegEM is to enable efficient circuit reconstruction in fully stained 3D-EM datasets by combination of skeleton reconstructions with automated volume segmentations. For this, it was crucial to make 3D-EM classifiers sufficiently reliable such that they can be combined into full-neuron reconstructions by assembly along the center line of neurites as shown above.

In addition to this concrete application, it was however desirable to compare the SegEM volume segmentations to the performance of other segmentation benchmarks (Figure 7). For this, we computed voxel, rand, and warping error using the routines made available in the ISBI EM challenge (see Supplemental Experimental Procedures; Figure 7H). SegEM results according



**Figure 6. SegEM Circuit Reconstruction Time Estimates**

(A) Projected circuit reconstruction cost measured in required work hour investment for two example circuits. Top left, circuit within layer 4 of cerebral cortex comprising thalamocortical axons (TC) and postsynaptic spiny neurons (L4 ss). Reconstructing a circuit of 30 TC axons and 30 spiny neurons would consume about 2,160 work hours for neurite tracing (at 2- to 4-fold tracing redundancy) and about 50 hr for synapse validation. Even a circuit of 100 axons and 100 dendrites is realistic (total effort: 7,700 work hours, i.e., less than \$80,000 salary investment). Bottom right, examples for circuits between bipolar neurons and amacrine cells in mouse retina. Note that for novel dataset, an additional volume labeling investment of 1,500–2,000 hr may be required. See [Supplemental Experimental Procedures](#) for details of calculations.

(B) SegEM analysis procedure for large-scale 3D-EM datasets, exemplified by the analysis timeline for a SBEM dataset from mouse somatosensory cortex, where a volume of  $(450 \mu\text{m})^3$  contains layer 4 of one cortical column and the required voxel size is about  $(12 \times 12 \times 25 \text{ nm})^3$  (dataset size  $\sim 25 \text{ TB}$ ). All numbers are approximate. SBEM: SBEM imaging (60 days) assumes an effective imaging speed of  $\sim 5 \times 10^6 \text{ vx/s}$  (including cutting, motor movements, etc.). Vol.: volume training data annotation for CNN training;  $\sim 200$  dense volume labelings of  $(100 \text{ vx})^3$  consuming  $\sim 1,600$  work hours (wh) distributed between 30–40 students over 2 weeks. Skel.: skeleton training data annotation for segmentation optimization; 5 students,  $\sim 300$  work hours (wh) over 2 weeks. CNN train.: classifier training/selection procedure (see [Figure 4](#)) over approximately 28 days (28 selection iterations as in [Figure 4B](#)). Segm. optim.: segmentation optimization ([Table 2](#)) using skeleton training data; about 2–5 days of computation. Alignment: alignment of SBEM images and conversion to Knossos data format (3D). Preproc.: dataset preprocessing (after entire dataset acquired); gray scale equalization; blood vessel masking, nucleus masking, potential correction of tiling effects (about 7 days of computation). Classification: dataset classification (forward pass of best CNN from selection procedure): 77 days on 56-GPU cluster. Segm.: local dataset segmentation in  $(128 \text{ vx})^3$  cubes with 10 vx margin on all sides on a 70–80-CPU cluster. Skeleton Circ. Rec: skeleton-based circuit reconstruction (at 4–8 hr per mm path length). Syn. Val.: synapse validation (at about 0.5–1 min per contact).

to the ISBI challenge metrics were among the best with respect to the warping error (aimed at directly optimizing topology of neurites) and rand error ([Figure 7I](#)). This is notable since the image data used in our (larger-volume) SegEM challenge was at resolution about 10-fold lower than that in the ISBI small-volume challenge (see [Figure 7H](#), left panels).

However, such comparisons should be treated with caution, since the ISBI challenge is evaluated only on 2D-EM data slices ([Figure 7H](#)), and voxel based metrics can be highly biased by the concrete shape and size of local objects contained in the test volume.

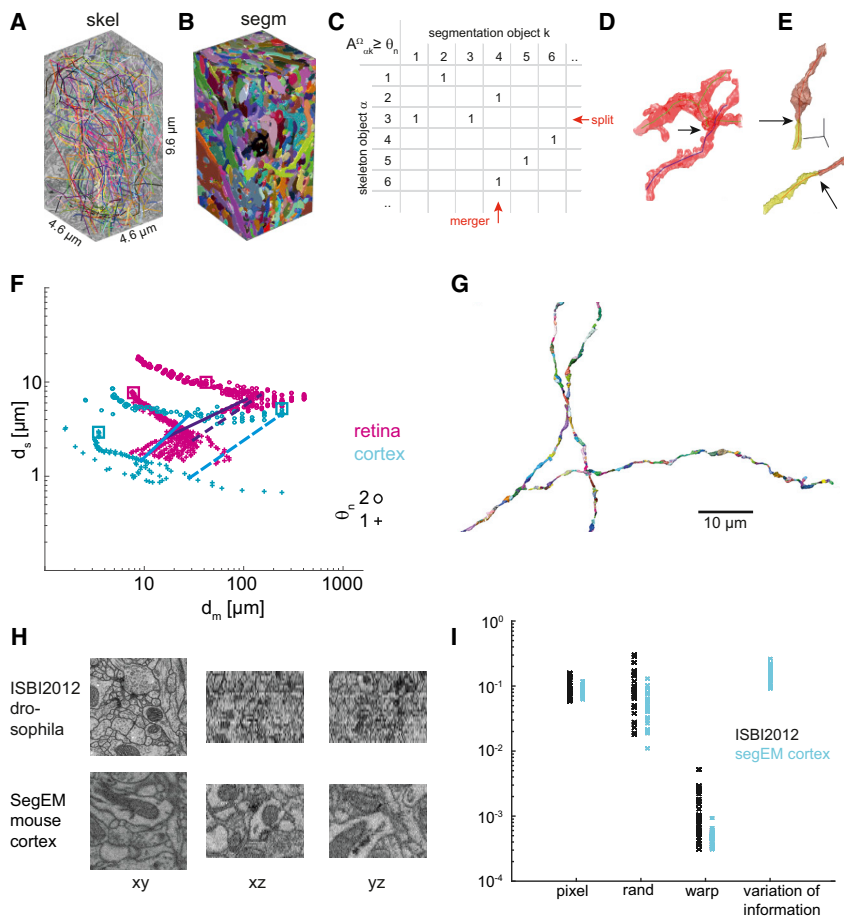
We therefore used an additional metric to measure SegEM segmentation error rates: we compared the volume segmentations to manually annotated skeleton (center-line) neurite reconstructions ([Figures 7A–7G](#)). This metric (skeleton-based split-merger metric; [Figures 7C–7F](#)) evaluates neurite continuity along the main axis of the neurites. Since the center line path length of neurons is on the scale of millimeters, such a metric therefore evaluates the most challenging aspect of neurite reconstructions: volume pieces as small as 50 nm in diameter, but at millimeters path distance, have to be correctly assigned to each other. The skeleton-based split-merger metric measures the rate of splits, i.e., breaks in the volume segmentation along the neurite center line axis ([Figure 7E](#)), and the rate of mergers, i.e., volume segmentation objects that link together two separate

neurite skeletons ([Figure 7D](#); see [Experimental Procedures](#) for details of the calculation or evaluateSeg.m in [Data S8](#) or SegEM package on github). Both measures are expressed as their inverse, i.e., the average distance between splits  $d_s$  and the average distance between mergers  $d_m$ .

[Figure 7F](#) shows the split-merger distances in the retina dataset for 372 combinations of segmentation parameters using CNN<sub>3,2</sub> and for 99 combinations of segmentation parameters using CNN 20130516T204040<sub>8,3</sub> for cortex. Error distances are reported for two node thresholds  $\theta_n$  indicating the minimum number of skeleton nodes that had to overlap with a segmentation object to be considered a match; higher values of  $\theta_n$  result in more resistance to noise from the imprecise manual placement of skeleton nodes.

The optimum inter-error distance  $1/(1/d_s + 1/d_m)$  for  $\theta_n = 1, 2$  was 3.9  $\mu\text{m}$ , 7.9  $\mu\text{m}$ , respectively (retina), and 1.6  $\mu\text{m}$ , 4.9  $\mu\text{m}$  (cortex; see [Table 2](#) for the corresponding optimal segmentation parameter sets). This optimum inter-error distance assumes splits and mergers in the automated segmentation to be of equally detrimental effect.

However, the key goal of our automated segmentation was to enable the efficient combination of skeleton reconstructions and volume segmentations for whole-neuron reconstructions. For this application, the optimal segmentation is one with maximal merger distances under the constraint that objects



**Figure 7. Segmentation Metrics and SegEM 3D Segmentation Challenge for Connectomics**

(A–C) Quantification of segmentation accuracy by comparison of automated volume segmentation  $\Omega$  (shown in B, classifier CNN3,2, see Table 2;  $r_{se} = 1$ ,  $\theta_{mg} = 0.3$ ,  $\theta_{ms} = 150$ ) with dense human-labeled skeleton annotation (A) yields overlap matrix  $A_{\alpha k}^{\Omega}$  between skeleton objects  $\alpha$  and segmentation objects  $k$ , which is binarized at node-overlap threshold  $\theta_n$  (C, see Experimental Procedures). Row sums exceeding 1 integrate to total split count; column sums exceeding 1 integrate to total merger count.

(D and E) Examples of mergers (D) and splits (E) (see Data S4 and S5 for complete gallery).

(F) Skeleton-based split-merger metric: average distance between splits ( $d_s$ ) and mergers ( $d_m$ ) for 372 segmentations obtained from varying the segmentation parameters as shown in Figure 2D for retina (magenta) and 99 segmentations cortex (cyan). Error distances are reported for skeleton-node-to-segmentation object overlap thresholds  $\theta_n = 1, 2$  (the larger threshold is less sensitive to node placement noise; see Experimental Procedures). Squares: optimal inter-error distance for  $\theta_n = 1, 2$  (3.85  $\mu m$ , 7.91  $\mu m$  for retina and 1.58, 4.93  $\mu m$  for cortex, respectively). Dashed lines: segmentation used in Figures 4 (retina, purple line) and 5 (cortex, blue line); see also (G); solid lines: segmentation used in all other figures (lines connect split-merger point for  $\theta_n = 1, 2$ ).

(G) Cortex axons and overlapping segmentation objects from segmentation in (F) (blue dashed line) to illustrate average object size (2 axons from Figure 5).

(H) Comparison of image data from ISBI 2012 2D-EM challenge and SegEM 3D-EM challenge. Note

higher resolution of data in xy (ISBI) versus problematic resolution and alignment in third dimension. Shown data are the entire training dataset for ISBI challenge. SegEM challenge comprises 279 of the volumes shown in (H).

(I) Comparison of all results submitted to ISBI 2012 2D-EM segmentation challenge as reported on [http://brainiac2.mit.edu/isbi\\_challenge/](http://brainiac2.mit.edu/isbi_challenge/) (black) and SegEM performance on test set of 34 (100 voxel)<sup>3</sup> regions from cortex dataset (blue) evaluated using the ISBI metrics (see Supplemental Experimental Procedures for details of calculation). Scale bars, 1  $\mu m$  in (D) and (E); 10  $\mu m$  in (G).

are rarely split longitudinally (example of longitudinal split: Figure S2C). Under this constraint, we find at node threshold  $\theta_n = 2$  an inter-merger distance of 151.3  $\mu m$  and inter-split distance of 7.4  $\mu m$  (Figure 7F, purple dashed line) to be most useful for whole-cell reconstruction in retina. For cortex, the segmentation used for whole-cell reconstruction had an inter-merger distance of 241.5  $\mu m$  and inter-split distance of 4.6  $\mu m$  (Figure 7F, blue dashed line; see Figure 7G for an example skeleton-segmentation overlay of a cortical axon using this segmentation setting).

Thus, in addition to enabling high-throughput circuit reconstruction, SegEM provides all training data and test metrics for large, truly 3D-EM datasets. The supplied training data comprise 987  $\mu m^3$  of labeled data, thus about 100 times more than that in the available EM challenges. Importantly, these data are sampled from many locations within a large 3D-EM dataset, providing properly generalizing classifiers. With this, SegEM also becomes a next-generation benchmark for machine learning in connectomics.

## DISCUSSION

We have developed SegEM, a semi-automated volume segmentation toolset for circuit reconstruction in fully stained 3D-EM image data (Figure 6). We applied our toolset to neuronal tissue from mouse retina and cerebral cortex. The exemplary synaptic innervation analyses (Figures 4 and 5) illustrate that SegEM resolves the tradeoff between synapse detectability in fully stained EM data and the reconstruction efficiency gain obtained by crowd-sourced skeleton reconstructions, which is required for dense circuit reconstruction in large-scale EM data.

We have proven the applicability to two types of neuronal tissue data even though the voxel-based classifiers do not generalize well between volume datasets from different neuronal tissues (Figures S2D and S2E). This may point to relevant local geometrical differences between peripheral and central nervous tissue.

While throughput of image analysis is still the major bottleneck in high-resolution connectomics, SegEM has made a substantial



difference in our own circuit analyses, reducing the investment per circuit analysis project by greater than 10-fold.

Further improvements in automated segmentation and human-machine interaction will be required for the most ambitious projects, such as whole-brain circuit reconstruction in the farther future, but SegEM has the ambition to make high-resolution connectomics a ready-to-use technique for many laboratories today (Figures 4, 5, and 6) and to furthermore provide a benchmark for fully 3D-EM-based connectomic data analysis (Figure 7).

## EXPERIMENTAL PROCEDURES

For the application of SegEM, please follow the detailed instructions in the [Supplemental Experimental Procedures](#).

### 3D-EM Image Datasets

The “retina” EM dataset (designated ek563, from [Briggman et al., 2011](#) and [Helmstaedter et al., 2013](#)) was acquired using serial blockface EM (SBEM, [Denk and Horstmann, 2004](#)). Tissue from a P30 mouse retina was stained using a conventional en bloc EM protocol comprising osmium tetroxide, thiocarbonylhydrazide amplification, and uranyl acetate staining steps (for details, see [Briggman et al., 2011](#)). SBEM images were aligned as described in [Briggman et al. \(2011\)](#): briefly, cross-correlation-derived shift vectors were computed in overlapping image regions, shift vectors were globally least-square optimized, and images were shifted using fourier-shift interpolation. The image data ( $4,096 \times 5,304 \times 5,760$  voxels; voxel size  $12 \times 12 \times 25$  nm<sup>3</sup>; dataset size  $50 \times 65 \times 145$  μm<sup>3</sup>; [Briggman et al., 2011](#)) were normalized to zero mean and unit SD on a grid with spacing of  $(128 \times 128 \times 128)$  vx<sup>3</sup> and cubes of size  $(163 \times 163 \times 143)$  vx<sup>3</sup> for the whole dataset. Each training volume was accordingly normalized (training volumes varied in size from  $[256 \times 256 \times 256]$  vx<sup>3</sup> to  $[512 \times 512 \times 256]$  vx<sup>3</sup>; [Table S1](#)).

The “cortex” EM dataset (designated 2012-09-28\_ex145\_07x2; K.M.B. and M.H., unpublished data) was acquired using SBEM. Tissue from S1 cortex of a P28 mouse was stained using a conventional en bloc EM protocol similar to the retina dataset. Image data ( $8,274 \times 5,338 \times 3,321$  voxels; voxel size  $11.24 \times 11.24 \times 28$  nm<sup>3</sup>; dataset size  $93 \times 60 \times 93$  μm<sup>3</sup>) were normalized to approximately zero mean (unnormalized mean: 122) and unit SD (unnormalized SD: 22) for the whole dataset (see [Figure S1H](#)). Each training volume was accordingly normalized (size of training volumes:  $[100 \times 100 \times 100]$  vx<sup>3</sup>). Images were aligned as for ek0563, but shift vectors between images were weighted before global optimization. Note that a re-evaluation of effective cutting thickness indicated a voxel size of  $11.24 \times 11.24 \times 26$  nm<sup>3</sup>; this correction was not applied to the data in this paper. All procedures were approved by the local animal care committee and were in accordance with the law of animal experimentation issued by the German Federal Government.

### Split and Merger Rates

The densely skeletonized segmentation test volume (see above and [Figure 7A](#)) was used to calculate the average length between merger ([Figure 2C](#)) and split ([Figure 2D](#)) errors ([Turaga et al., 2010](#)). To allow comparison of split-merger metrics between independently skeletonized datasets of different neurite geometry (training and test skeletonizations from retina and cortex), we first equilibrated skeleton node densities between skeleton sets (resulting average inter-node distance: 460 nm).

Then, we calculated a skeleton node-to-segmentation object overlap matrix  $A_{\alpha k}^{\Omega}$  reporting for each combination of skeleton  $\alpha$  and segmentation object  $k$  (in segmentation  $\Omega$ ) the number of skeleton nodes in  $\alpha$  that overlapped with any voxel labeled as segmentation object  $k$  (see [Figure 7C](#)):

$$A_{\alpha k}^{\Omega} = \sum_{m=1..N_{\alpha}} \delta(\Omega(i(\alpha_m)), k),$$

with  $N_{\alpha}$  the total number of nodes in skeleton  $\alpha$ ,  $i(\alpha_m)$  the voxel location of the  $m^{\text{th}}$  node in skeleton  $\alpha$ , and  $\Omega(i)$  the label of segmentation  $\Omega$  at location  $i$ . This

matrix was then binarized at a node threshold  $\theta_n$ , which was varied between 1 and 2 to assess the influence of labeling (i.e., skeleton node placement) noise

$$(\text{see Figure 7F}): A_{\alpha k}^{\Omega\#} = \begin{cases} 1 & \text{if } A_{\alpha k}^{\Omega} \geq \theta_n \\ 0 & \text{otherwise} \end{cases}.$$

The number of mergers and splits was then calculated as

$$n_{\text{merger}} = \sum_{\alpha=1..N_S} \left( \left( \sum_{k=1..N_{\Omega}} A_{\alpha k}^{\Omega\#} - 1 \right) * \Theta \left( -1.5 + \sum_{k=1..N_{\Omega}} A_{\alpha k}^{\Omega\#} \right) \right)$$

and

$$n_{\text{splits}} = \sum_{\alpha=1..N_{\Omega}} \left( \left( \sum_{k=1..N_S} A_{\alpha k}^{\Omega\#} - 1 \right) * \Theta \left( -1.5 + \sum_{k=1..N_S} A_{\alpha k}^{\Omega\#} \right) \right),$$

with  $N_S$  and  $N_{\Omega}$  as the number of skeletons and segments, respectively, while  $\Theta$  is the Heaviside step function. Finally, the average distance between splits  $d_s$  and between mergers  $d_m$  was calculated as  $d_s = L / n_{\text{split}}$  and  $d_m = L / n_{\text{merger}}$ , with  $L$  the total skeleton path length in the segmentation test volume ( $L = 0.48$  mm and  $1.21$  mm for cortex and retina test volumes, respectively). In cases where no split or no merger was detected,  $n_{\text{merger}}$  or  $n_{\text{split}}$  were set to 1 to limit error distance confidence. Segmentations resulting in only 1 merger or only 1 split were not evaluated. Optimal inter-error distance (the minimum of  $1/(d_s + 1/d_m)$ ) was determined by a sliding 3-nearest neighbor average in the ds-dm plane ([Figure 7F](#)).

Note that to avoid artifacts by different fractions of wall voxels in different segmentations, the split-merger metric should be computed on fully grown segmentations only (no remaining wall classified voxels; [Figure 7F](#) was computed for such fully grown out segmentations only).

The code for computing this metric is contained in `evaluateSeg.m` ([Data S8](#)).

### SegEM 3D Image Segmentation Challenge for Connectomics

For participation in the SegEM 3D image segmentation challenge for connectomics, please follow the instructions below. SegEM ([segem.io](#)) provides 279 densely volume-annotated volumes in which  $(100 \text{ voxel})^3$  are each labeled. These volumes are sampled from throughout the cortex dataset, providing a representative sampling of neurite geometry and image statistics (see [Table S2](#) for a detailed overview). To allow training of classifiers of larger field of view, we supply the raw data for each labeled volume of size  $200 \times 200 \times 150$  voxel, i.e., with a border of  $100 \times 100 \times 50$  voxel around the labeled volume.

For testing the split-merger metric, we provide a densely skeletonized test set ([Data S3](#)), which was used for [Figure 7F](#). Split-merger metrics calculation is contained in `cortex/segmentation/evaluateSeg.m` of the SegEM package ([Data S8](#) or [segem.io](#)). For calculation of pixel, warp, rand error, and variation of information as in [Figure 7I](#), please use the code provided by the ISBI 2012 2D-EM segmentation challenge (see above).

For submission of results to the SegEM 3D image segmentation challenge, an email containing the code to process raw data volumes of the same size as the training data should be sent to [segEMchallenge@brain.mpg.de](mailto:segEMchallenge@brain.mpg.de). This will be evaluated on 34 test stacks of the same size as the training stacks given above and on an additional densely skeletonized test set from several regions in the cortex dataset. Random examples of the raw data of these hidden test sets will be provided on [segem.io](#), but not the respective labels. Ranking of segmentation results will be continuously updated on [segem.io](#) with respect to the ISBI challenge metrics and skeleton-based split-merger metrics.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, two tables, and nine data files and can be found with this article online at <http://dx.doi.org/10.1016/j.neuron.2015.09.003>.

## AUTHOR CONTRIBUTIONS

All work was performed by M.B. and M.H. K.M.B. acquired the cortex dataset.

## ACKNOWLEDGMENTS

We thank Viren Jain, Srinivas Turaga, and Sebastian Seung for fruitful discussions in the early phase of the project; Sarah Kaspar for initial classifier training; and Martin Zauser for contributions to code. We thank Yates Buckley, Winfried Denk, Emmanuel Klinger, Helene Schmidt, and Jakob Straehle for comments on earlier versions of the manuscript and Chris Roome and Mario Hilpert for support of the compute infrastructure. We are especially grateful to Heiko Wissler, Fabian Isensee, and Tim Decker for their excellent help with annotation management, synapse annotation, and cell-type classification. We thank Sophie Dittmer, Marco Reiner, Lisa Geiser, Katharina Haase, Bianca Hauber, Julian Sieber, Andre Antunes, Stephanie Best, Simone Pfarr, Claudia Klein, Christina Fianke, Jenny Bauer, Philipp Bastians, Arlie Zegarra, Julia Volz, Patrick Weber, Maximilian Schiedeck, Stephanie Freiss, Sadia Oumohand, Annika Gable, Janika Briegel, Maria Mueller, Emilia Wiegand, Jessica Dietrich, Julia Pesch, Florian Viehweger, Torben Tannig, Niclas Rieger, Patrick Hofmann, Kathrin Schramm, Michael Kirchberger, Julia Hammerich, Stefanie Ehrhardt, Marie Reitz, Erlina Rachmad, Jenny Raetzer, Maximilian Scheller, Matthias Pohrath, Maret Heumannskaemper, Jeannie Schied, Christian Orlik, Lili Hocke, Corinna Eck, Kristina Ernst, Ralf Vogel, Thomas Schackel, Yannick Soehngen, Maike Schramm, Kai Fischer, Edem Atsiatorne, Julia Ricken, Ramtin Lichtenberger, and Martina Jonczyk for neurite skeletonization and Anikó Pálfi, Anna Buntjer, Anna Leinhos, Anna Rommerskirchen, Benedikt Martin, Carina Meyer, Carolin Willburger, Christian Röhrig, Dirk Ollech, Franziska Hentzschel, Hanna Jakobi, Hanna Rogg, Jana Munstermann, Johanna Lott, Josephine Reinhardt, Karin Bretzel, Katharina Kappler, Kathrin Haase, Kathrin Schleich, Kira Garbe, Lisa Keitel, Madeleine Joel, Marcel Smykalla, Martin Moll, Monika Berberich, Nina Wommelsdorf, Patricia Hegemann, Patrick Weber, Philip Gallandi, Raphael Foitzik, Rebekka Wiggers, Sarah Kaspar, Steffen Klein, Stephanie Best, Susanne Dettmer, Susanne Schuster, Victoria Wissdorf, and Wolfram Volkwein for volume tracing. M.H. is a shared stakeholder of the patent "Method and apparatus for image processing," Published Patent Application No. 20100183217.

Received: March 11, 2015

Revised: August 3, 2015

Accepted: August 27, 2015

Published: September 23, 2015

## REFERENCES

- Andres, B., Koethe, U., Kroeger, T., Helmstaedter, M., Briggman, K.L., Denk, W., and Hamprecht, F.A. (2012a). 3D segmentation of SBFSEM images of neuropil by a graphical model over supervoxel boundaries. *Med. Image Anal.* **16**, 796–805.
- Andres, B., Kroeger, T., Briggman, K., Denk, W., Korogod, N., Knott, G., Koethe, U., and Hamprecht, F. (2012b). Globally Optimal Closed-Surface Segmentation for Connectomics. In *Computer Vision – ECCV 2012*, A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, and C. Schmid, eds. (Springer Berlin Heidelberg).
- Bergstra, J., Yamins, D., and Cox, D.D. Hyperopt: A Python library for optimizing the hyperparameters of machine learning algorithms. *Proceedings of the 12th Python in Science Conference*, 2013. 13–20.
- Briggman, K.L., and Bock, D.D. (2012). Volume electron microscopy for neuronal circuit reconstruction. *Curr. Opin. Neurobiol.* **22**, 154–161.
- Briggman, K.L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature* **471**, 183–188.
- Cardona, A., Saalfeld, S., Schindelin, J., Arganda-Carreras, I., Preibisch, S., Longair, M., Tomancak, P., Hartenstein, V., and Douglas, R.J. (2012). TrakEM2 software for neural circuit reconstruction. *PLoS ONE* **7**, e38011.
- Ciresan, D., Giusti, A., Gambardella, L.M., and Schmidhuber, J. (2012). Deep neural networks segment neuronal membranes in electron microscopy images. *Adv. Neural Inf. Process. Syst.* **25**, 2843–2851.
- Denk, W., and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol.* **2**, e329.
- Denk, W., Briggman, K.L., and Helmstaedter, M. (2012). Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nat. Rev. Neurosci.* **13**, 351–358.
- Funke, J., Andres, B., Hamprecht, F.A., Cardona, A., and Cook, M. Efficient automatic 3D-reconstruction of branching neurons from EM data. *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, 16–21 June 2012 2012. 1004–1011.
- Hayworth, K., Kasthuri, N., Schalek, R., and Lichtman, J. (2006). Automating the collection of ultrathin serial sections for large volume TEM reconstructions. *Microsc. Microanal.* **12**, 86–87.
- Helmstaedter, M. (2013). Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nat. Methods* **10**, 501–507.
- Helmstaedter, M., Briggman, K.L., and Denk, W. (2011). High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nat. Neurosci.* **14**, 1081–1088.
- Helmstaedter, M., Briggman, K.L., Turaga, S.C., Jain, V., Seung, H.S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature* **500**, 168–174.
- Jain, V., Murray, J.F., Roth, F., Turaga, S., Zhigulin, V., Briggman, K.L., Helmstaedter, M.N., Denk, W., and Seung, H.S. Supervised learning of image restoration with convolutional networks. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007. IEEE, 1–8.
- Jain, V., Bollmann, B., Richardson, M., Berger, D.R., Helmstaedter, M.N., Briggman, K.L., Denk, W., Bowden, J.B., Mendenhall, J.M., and Abraham, W.C. Boundary learning by optimization with topological constraints. *Computer Vision and Pattern Recognition (CVPR)*, (2010a) IEEE Conference on, 2010. IEEE, 2488–2495.
- Jain, V., Seung, H.S., and Turaga, S.C. (2010b). Machines that learn to segment images: a crucial technology for connectomics. *Curr. Opin. Neurobiol.* **20**, 653–666.
- Jain, V., Turaga, S.C., Briggman, K.L., Helmstaedter, M.N., Denk, W., and Seung, H.S. (2011). Learning to Agglomerate Superpixel Hierarchies. *Adv. Neural Inf. Process. Syst.* **24**, 2.
- Jones, C., Liu, T., Cohan, N.W., Ellisman, M., and Tasdizen, T. (2015). Efficient semi-automatic 3D segmentation for neuron tracing in electron microscopy images. *J. Neurosci. Methods* **246**, 13–21.
- Kasthuri, N., Hayworth, K.J., Berger, D.R., Schalek, R.L., Conchello, J.A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T.R., et al. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell* **162**, 648–661.
- Kaynig, V., Vázquez-Reina, A., Knowles-Barley, S., Roberts, M., Jones, T.R., Kasthuri, N., Miller, E., Lichtman, J., and Pfister, H. (2015). Large-scale automatic reconstruction of neuronal processes from electron microscopy images. *Med. Image Anal.* **22**, 77–88.
- Knott, G., Marchman, H., Wall, D., and Lich, B. (2008). Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J. Neurosci.* **28**, 2959–2964.
- Lichtman, J.W., and Denk, W. (2011). The big and the small: challenges of imaging the brain's circuits. *Science* **334**, 618–623.
- Liu, T., Jones, C., Seyedhosseini, M., and Tasdizen, T. (2014). A modular hierarchical approach to 3D electron microscopy image segmentation. *J. Neurosci. Methods* **226**, 88–102.
- Nunez-Iglesias, J., Kennedy, R., Plaza, S.M., Chakraborty, A., and Katz, W.T. (2014). Graph-based active learning of agglomeration (GALA): a Python library to segment 2D and 3D neuroimages. *Front. Neuroinform.* **8**, 34.
- Saalfeld, S., Cardona, A., Hartenstein, V., and Tomančák, P. (2009). CATMAID: collaborative annotation toolkit for massive amounts of image data. *Bioinformatics* **25**, 1984–1986.
- Seyedhosseini, M., Kumar, R., Jurrus, E., Giuly, R., Ellisman, M., Pfister, H., and Tasdizen, T. (2011). Detection of neuron membranes in electron



- microscopy images using multi-scale context and radon-like features. *Med Image Comput Comput Assist Interv* 14, 670–677.
- Seyedhosseini, M., Sajjadi, M., and Tasdizen, T. Image segmentation with cascaded hierarchical models and logistic disjunctive normal networks. *Computer Vision (ICCV)*, 2013 IEEE International Conference on, 2013. IEEE, 2168–2175.
- Snoek, J., Larochelle, H., and Adams, R.P. (2012). Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* 25, 2951–2959.
- Sommer, C., Straehle, C., Kothe, U., and Hamprecht, F.A. Ilastik: Interactive learning and segmentation toolkit. *Biomedical Imaging: From Nano to Macro*, 2011 IEEE International Symposium on, March 30 2011–April 2 2011 2011. 230–233.
- Takemura, S.-Y., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P.K., Katz, W.T., Olbris, D.J., Plaza, S.M., Winston, P., et al. (2013). A visual motion detection circuit suggested by *Drosophila* connectomics. *Nature* 500, 175–181.
- Turaga, S.C., Briggman, K.L., Helmstaedter, M., Denk, W., and Seung, H.S. 2009. Maximin affinity learning of image segmentation. *arXiv preprint arXiv:0911.5372*.
- Turaga, S.C., Murray, J.F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., and Seung, H.S. (2010). Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural Comput.* 22, 511–538.
- Vazquez-Reina, A., Gelbart, M., Huang, D., Lichtman, J., Miller, E., and Pfister, H. Segmentation fusion for connectomics. *Computer Vision (ICCV)*, 2011 IEEE International Conference on, 2011. IEEE, 177–184.



## 6.2 webKnossos: Efficient online 3D data annotation for connectomics

**Abstract** We report webKnossos, an in-browser annotation tool for 3D electron microscopic data. webKnossos provides flight mode, a single-view egocentric reconstruction method enabling trained annotator crowds to reconstruct at a speed of  $1.5 \pm 0.6$  mm/h for axons and  $2.1 \pm 0.9$  mm/h for dendrites in 3D electron microscopic data from mammalian cortex. webKnossos accelerates neurite reconstruction for connectomics by 4- to 13-fold compared with current state-of-the-art tools, thus extending the range of connectomes that can realistically be mapped in the future.

**Contributions** M.H. initiated and supervised the project; K.M.B., M.B., T.B., N.R., T.W. and M.H. developed specifications and conceptual design with contributions by H.W.; T.B., D.B., J.F., T.H., P.O., N.R., T.W., D.W., G.W. and K.M.B. implemented the software; H.W., M.B., K.M.B. and F.D. provided data; K.M.B., M.H., H.W. and M.B. analyzed the data; M.H., K.M.B. and M.B. wrote the manuscript with contributions by all authors.

**Copyright** Reprinted by permission from Macmillan Publishers Ltd: Nature methods [Boergens et al., 2017], copyright 2017.

# webKnossos: efficient online 3D data annotation for connectomics

Kevin M Boergens<sup>1,3</sup> , Manuel Berning<sup>1,3</sup>, Tom Bocklisch<sup>2</sup>, Dominic Bräunlein<sup>2</sup>, Florian Drawitsch<sup>1</sup>, Johannes Frohnhofen<sup>2</sup>, Tom Herold<sup>2</sup>, Philipp Otto<sup>2</sup>, Norman Rzepka<sup>2</sup>, Thomas Werkmeister<sup>2</sup>, Daniel Werner<sup>2</sup>, Georg Wiese<sup>2</sup>, Heiko Wissler<sup>1</sup> & Moritz Helmstaedter<sup>1</sup> 

**We report webKnossos, an in-browser annotation tool for 3D electron microscopic data. webKnossos provides flight mode, a single-view egocentric reconstruction method enabling trained annotator crowds to reconstruct at a speed of  $1.5 \pm 0.6$  mm/h for axons and  $2.1 \pm 0.9$  mm/h for dendrites in 3D electron microscopic data from mammalian cortex. webKnossos accelerates neurite reconstruction for connectomics by 4- to 13-fold compared with current state-of-the-art tools, thus extending the range of connectomes that can realistically be mapped in the future.**

With the acceleration of 3D electron microscopic (EM) imaging of brain tissue<sup>1–3</sup>, image data sets sized tens of terabytes (TB) or even petabytes (PB) are becoming available. A cubic millimeter imaged at  $(15 \text{ nm})^3$  voxel (vx) size corresponds to 0.3 PB of data (Fig. 1a); a mouse brain imaged at the same resolution corresponds to 110 PB of data. Single neurons typically extend over a large fraction of the data set (Fig. 1a), making it impracticable to distribute data on hard drives to large numbers of annotators who want to follow the processes of entire neurons. At the same time, data analysis in connectomics is limited by the amount of human annotation time that can be recruited for a given analysis project<sup>4</sup>. Thus, enabling efficient distributed 3D data annotation in PB-sized data sets, ideally in browser, is essential.

The existing in-browser annotation tool for connectomics, CATMAID<sup>5,6</sup>, uses efficient data storage and transmission in 2D image planes (comparable to Google Maps<sup>7</sup>), which are sequentially browsed (Fig. 1b). While this approach makes data viewing and annotation seamless in the plane of imaging, 3D neurite tracing is slowed down by the time required to progress to the subsequent image plane. Under ideal high-bandwidth, low-latency connectivity conditions (like those within research institutions), this approach yields a reconstruction speed of about 470  $\mu\text{m}/\text{h}$

when an expert follows an axon through 3D serial blockface EM (SBEM<sup>1</sup>) data from mouse cortex (Fig. 1c). Under less optimal bandwidth and latency conditions (as are often experienced by student annotators at home and in mobile settings), however, reconstruction speed drops to 59–88  $\mu\text{m}/\text{h}$  (measured for regular 3G connectivity and for transcontinental access, Fig. 1c).

webKnossos (Fig. 1d, <https://webknossos.org>) uses 3D data storage and transmission in small cubic packages of  $32^3$  vx (Fig. 1b). Cubic 3D image data storage using cubes of  $128^3$  vx is being employed in KNOSSOS<sup>8</sup> and pyKNOSSOS<sup>9,10</sup>, standalone data annotation applications for connectomics. We reduced 3D cube size to  $32^3$  vx for lag-free in-browser data transmission and for enabling the flight-mode data visualization introduced below. webKnossos enables data interaction in 3D EM images displayed in orthogonal planes (Fig. 1d) at a speed of about 2 mm/h (Fig. 1c), which drops to 0.7–1.2 mm/h for regular 3G connectivity and transcontinental annotation (Fig. 1c). Thus, 3D data visualization when following an axon is about 4-fold faster under ideal and up to 13-fold faster under nonoptimal connectivity settings than with existing in-browser tools (Fig. 1c).

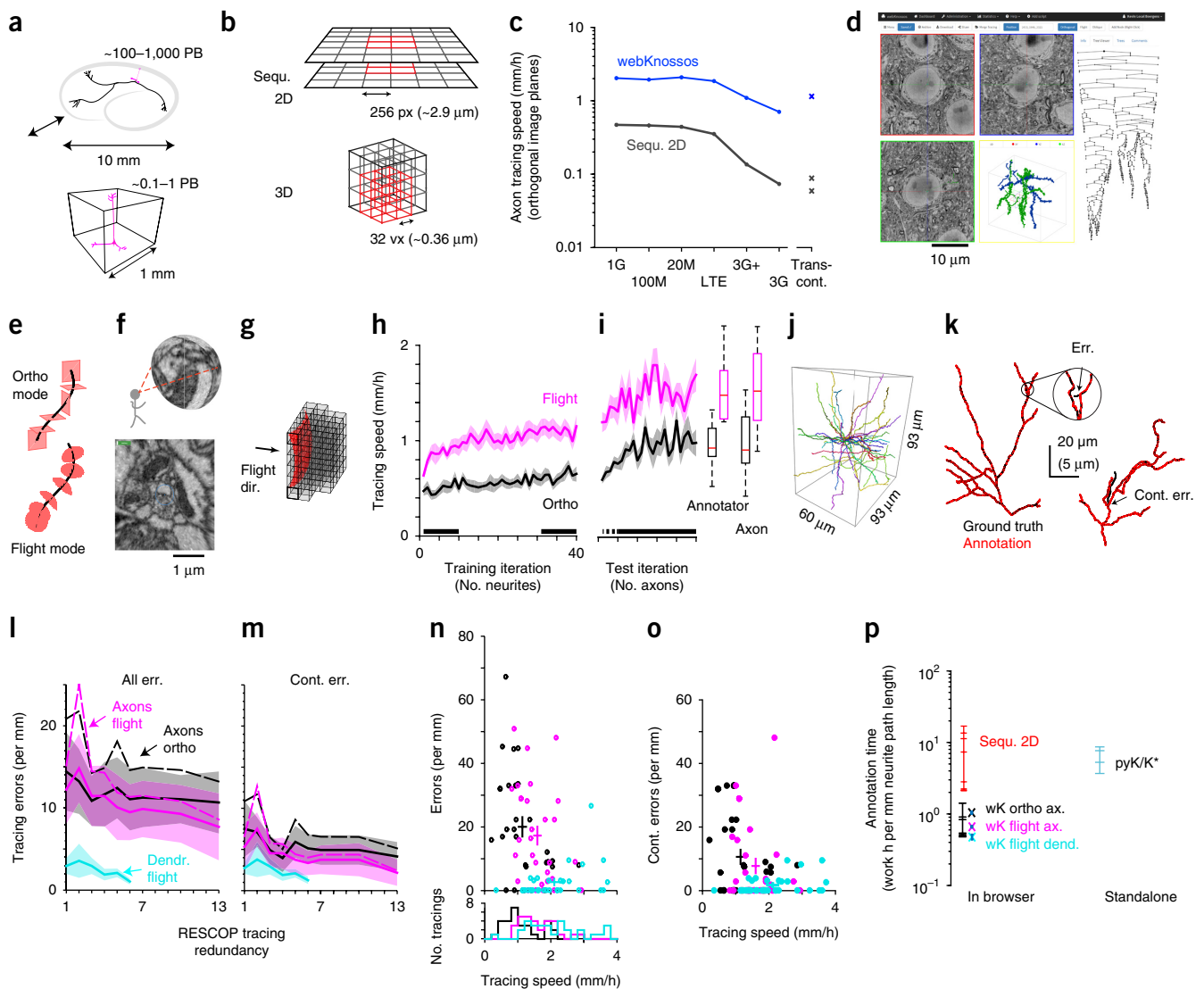
We next tested whether student annotators can be trained to interact with 3D brain image data at such speeds. Previously, annotators interacted with the image data using orthogonal image projections in the three cardinal planes for following the neuronal processes and for their annotation as ‘skeletons’ (Fig. 1d,e, ‘ortho mode’, KNOSSOS<sup>8</sup>). Effective tracing speed was 100–270  $\mu\text{m}/\text{h}$ <sup>4,8,10–12</sup> for reconstructions in mouse retina, zebrafish olfactory bulb and mouse cortex. We asked whether annotators can be trained to annotate faster in ortho mode, and whether a more intuitive data presentation can further accelerate human annotation. For the latter, we developed ‘flight mode’, in which the 3D image data are sampled on a hemisphere centered at the annotator’s current position (Fig. 1f). To enable such non-orthogonal data transmission and display in browser, we used (in addition to the small 3D cube size (Fig. 1b)) a simple form of path prediction (the data being loaded in a stump in direction of flight; Fig. 1g). Flight mode requires the EM image data to be sufficiently well aligned in 3D, as is routinely the case for neuronal tissue imaged using SBEM<sup>1</sup>. In flight mode, the annotator focuses on centering the target cursor onto the axon or dendrite being followed, steering the orientation with the mouse or keyboard while moving forward (Supplementary Video 1). We suspected that this focusing on one intuitive egocentric visualization and interaction may accelerate annotation, since the user does not have to explicitly recenter the viewport and switch image plane orientation for processes running in off-axis directions (Fig. 1e).

To investigate whether flight mode in fact accelerates human 3D image data annotation, we trained 51 student annotators on

<sup>1</sup>Department of Connectomics, Max Planck Institute for Brain Research, Frankfurt, Germany. <sup>2</sup>Scalable minds UG (haftungsbeschränkt) & Co. KG, Potsdam, Germany.

<sup>3</sup>These authors contributed equally to this work. Correspondence should be addressed to M.H. ([mh@brain.mpg.de](mailto:mh@brain.mpg.de)) or K.M.B. ([kevin.boergens@brain.mpg.de](mailto:kevin.boergens@brain.mpg.de)).

RECEIVED 23 MARCH 2016; ACCEPTED 17 MAY 2017; PUBLISHED ONLINE 12 JUNE 2017; DOI:10.1038/NMETH.4331

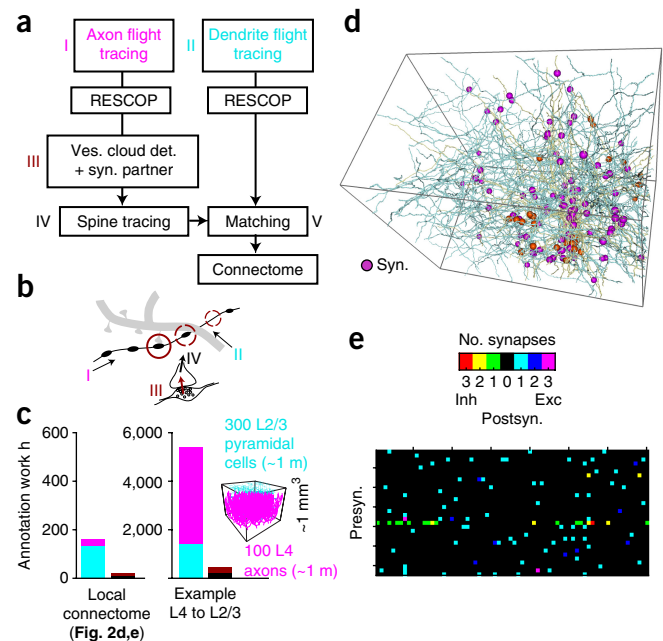


**Figure 1** | In-browser 3D annotation of axons and dendrites for connectomics. (a) Sketch of mouse whole-brain 3D EM data set size (top) and 1 mm<sup>3</sup> of cerebral cortex (bottom) compared with the extent of a typical single pyramidal neuron (dendrites, magenta; axon, black). (b) Sketch of online data delivery modes using lateral prefetching in 2D (top, Google Maps, CATMAID) and in 3D (bottom, webKnossos). (c) Comparison of data transmission speed when following a neurite using 2D image prefetching (black, in CATMAID) and 3D data prefetching in webKnossos (blue, ortho mode) under different bandwidth and latency conditions. Transcont, transcontinental access (<https://www.openconnectomeproject.org> from Europe, custom CATMAID and webKnossos from South America, bottom to top). (d) webKnossos in-browser user interface with orthogonal viewports (xy, yz, xz), one 3D skeleton viewport, and the abstract tree viewer (right). (e) Sketch of viewing surface orientation (red) in orthogonal mode (top) and flight mode (bottom). (f) Flight-mode egocentric 3D image sampling on a hemisphere (top), yielding a single flight-mode data view (bottom). (g) Example of 3D prefetching in flight mode given flight direction and current position (gray, prefetched webKnossos cubes; red, flight-mode image surface). (h) Annotator speed training in ortho mode (black,  $n = 25$  annotators) and flight mode (magenta,  $n = 26$  annotators) on 40 neurites in cortex (randomly ordered per annotator). (i) Tracing speed test on 20 randomly selected cortical axons (including branches) performed by 26 annotators 8 weeks after training. Dashed line, first five neurites presented again from training. Solid line, 20 test axons. Box plots report tracing speed for these 20 test axons in ortho (black) and flight mode (magenta), reported over  $n = 26$  annotators (left) and  $n = 20$  axons (right); flight  $1.51 \pm 0.04$  mm/h, ortho  $0.96 \pm 0.03$  mm/h ( $n = 520$ , mean  $\pm$  s.e.m.). (j) Display of 20 test axons within data set boundary. (k) Illustration of tracing error measurement in one of the 20 axons—local errors (inset, less than 10- $\mu$ m path length) and continuation errors (right). Black, ground truth; red, six-fold consolidated flight-mode tracing. (l) Tracing errors reported over tracing redundancy (using RESCOP<sup>8</sup>) for 10 axons in ortho mode, flight mode and for 10 dendrites in flight mode (cyan), respectively. (m) same as l but only continuation errors (see k). Dashed lines in l and m, path length corrected for each RESCOPed skeleton (see Online Methods). (n) Relation between tracing speed and error rates for single-annotator reconstructions ( $n = 30$ ) of axons and dendrites (colors as in l and m). Crosses indicate mean  $\pm$  s.e.m. (o) same as n for continuation errors only. (p) Summary comparison of annotation time requirements for neurite reconstruction. Data from c, (black, webKnossos, <https://webknossos.org>; red, CATMAID) and from i and n (crosses). Asterisk, annotation consumption documented in published work; pyK, consumption by experienced annotators<sup>10</sup>; K, KNOSSOS annotation consumption from refs. 8, 11 and 12.

40 neurites in a 3D EM data set from mouse cortex sized  $93 \times 60 \times 93 \mu\text{m}^3$ , imaged at  $11.24 \times 11.24 \times 28 \text{ nm}^3$  using SBEM (data not shown; **Fig. 1h**). The sequence of neurites to trace was shuffled for each student (total path length of all 40 neurites, 6.11 mm; number of branch points, 258; overall branch point rate, 42 per mm). The student annotators came from our pool of trained annotators and were thus experienced in connectomic data reconstruction (**Supplementary Fig. 1**). They were randomly assigned to two groups of 25 and 26 students, respectively. One group was asked to trace in ortho mode, the other in flight mode. We supplied both groups with an 8.5-min tutorial movie (a separate movie for each group), which encouraged them to increase their movement speed whenever they felt they could go faster during reconstructions (**Supplementary Fig. 1b,d**). To enable constantly maximized tracing speed for each annotator, we automatically tracked the fraction of time during which the annotators proceeded at the preset movement speed (i.e., when holding the forward key pressed while navigating) and reminded them to increase their preset movement speed if they spent more than 75% of their tracing time constantly pressing the forward key. In ortho mode (**Fig. 1h**), annotators initially traced at  $0.49 \pm 0.04 \text{ mm/h}$  (first ten processes traced, mean  $\pm$  s.e.m.) and accelerated to  $0.63 \pm 0.05 \text{ mm/h}$  (last ten processes traced,  $P < 10^{-3}$ , Wilcoxon signed-rank test). In flight mode (**Fig. 1h**), annotators started at a speed of  $0.84 \pm 0.05 \text{ mm/h}$  (faster than ortho mode,  $P < 10^{-4}$ , Wilcoxon rank-sum test) and were able to increase their speed to an average of  $1.11 \pm 0.07 \text{ mm/h}$  ( $P < 10^{-4}$ , Wilcoxon signed-rank test), 1.8 times faster than ortho mode tracing ( $P < 10^{-5}$ , Wilcoxon rank-sum test). These data indicate that a substantial reconstruction speed gain comes from per-user speed maximization, enabled by efficient 3D data handling, and an additional gain from the egocentric flight mode data interaction.

To test whether this tracing speed can be routinely achieved for axons in mammalian cerebral cortex (**Fig. 1i,j**), we next randomly selected 20 axons from a  $(2.5 \mu\text{m})^3$  region in the same data set (**Fig. 1j**) and asked students trained in ortho mode to again use ortho mode and students trained in flight mode to again use flight mode. In each group, 13 of the trained students participated in this second experiment 8 weeks after the training (their initial training performance had been indistinguishable from the whole group,  $P > 0.24$ , Wilcoxon rank-sum test). We first presented five of the processes reconstructed during training to calibrate the persistence of the training effects, and then we presented the 20 new axon seeds in random order to all tracers (**Fig. 1i**; note that the five neurites from the training session were not included in the final speed measurement). Tracers resumed annotation at the speed attained during training and were able to further accelerate, yielding a reconstruction speed of  $0.96 \pm 0.03 \text{ mm/h}$  in ortho and  $1.51 \pm 0.04 \text{ mm/h}$  in flight mode (**Fig. 1i**, mean  $\pm$  s.e.m.,  $n = 20$  previously unseen randomly selected cortical axons; total path length, 2.53 mm; overall branch-point rate, 39 per mm).

But were faster tracings more error prone? We next quantified the rate of errors for 10 randomly drawn axons out of the 20 test axons. For each axon, we manually counted the number of incorrect stops and incorrect continuations (**Fig. 1k**, performed by two expert annotators blinded to tracing mode, see Online Methods) and distinguished between errors yielding continuation mistakes (i.e., a premature stop or missed branch of a major part of the axon) or local errors (yielding less than  $10 \mu\text{m}$  neurite



**Figure 2** | Connectome reconstruction using webKnossos. (a) Flow chart and illustration (b) of connectome reconstruction steps. Ves. cloud det., vesicle cloud detection; syn. partner, identification of postsynaptic partner. (c) Comparison of annotation times for axons, dendrites, spines and synapses including the speed gains for neurite reconstruction in webKnossos (**Fig. 1**) for the local connectome shown in panels d and e and for a sparse example reconstruction (estimate, see inset, 100 layer-4 cortical axons innervating 300 layer-2/3 pyramidal cells within a cubic millimeter of cortex tissue, amounting to 2 m total path length). (d) Reconstruction of 497 dendrites and 32 axons in local SBEM data set from mouse cortex following the workflow in a. Colored spheres indicate excitatory (Exc., violet) and inhibitory (Inh., red) synaptic contacts (syn.). (e) Resulting connectome between 32 presynaptic axons (presyn.) and 497 postsynaptic (postsyn.) dendrites (only the 70 innervated dendrites are shown).

loss or neurite addition). **Figure 1l** reports the rate of errors for flight-mode and ortho-mode tracings for single-annotator reconstructions and for consolidations of multiple reconstructions of the same axon (consolidated using RESCOP<sup>8</sup>). We found that, first, the rate of tracing errors was not distinguishable between ortho and flight mode tracings ( $P > 0.34$  for all errors,  $P > 0.28$  for continuation errors, Wilcoxon rank-sum test). Second, the average error rate for single-annotator reconstructions obtained at the achieved speed in webKnossos was not worse than the error rates reported previously in mouse retina<sup>8,11</sup> and cortex<sup>13</sup>. The rate of continuation errors (**Fig. 1m**) was  $7.5 \pm 3.4$  per mm in ortho mode and  $5.3 \pm 3.0$  per mm in flight mode for single-annotator tracings. We finally asked whether a speed-accuracy tradeoff could be observed in either of the tracing modes. For this we correlated the rate of errors with tracing speed in single-annotator tracings (**Fig. 1n**). No positive correlation could be found for ortho mode nor for flight mode (ortho  $r = -0.5$ ,  $P = 0.007$ , flight  $r = -0.4$ ,  $P > 0.05$ , Pearson's correlation). This also was true when only analyzing the continuation errors (**Fig. 1o**, ortho  $r = -0.34$ ,  $P > 0.05$ , flight  $r = -0.20$ ,  $P > 0.28$ , Pearson's correlation).

We thus conclude that annotators can be trained to trace cortical axons at  $1.51 \pm 0.04 \text{ mm/h}$  in flight mode without a reduction in accuracy. webKnossos can support this speed online and



in browser, and it provides a 6- to 15-fold improvement over published tracing speeds, depending on the reconstructed data sets (summarized in Fig. 1p).

To determine tracing speeds for dendrites, we reconstructed the shafts of ten randomly chosen dendrites (1.8-mm total path length, overall branch point rate 16 per mm) and measured tracing errors as described above. Since dendrites are about three times wider in diameter<sup>14</sup>, annotators could zoom out further and fly along dendrites faster than they do along axons. We found that tracing speed was  $2.11 \pm 0.16$  mm/h including branch-point reconstruction, and single-annotator error rates were  $2.7 \pm 0.69$  errors per mm dendrite (Fig. 1l–o).

In mammalian brains, which constitute a main challenge of connectomics, about 90% of the neuronal processes are axons<sup>14,15</sup>. The speed gain for representative axon reconstruction in flight mode was therefore critical for the acceleration of connectomic reconstruction in mammalian cortex. However, the reconstruction of connectomes additionally requires the identification of synapses and the assignment of postsynaptic partners to the respective neuronal cell bodies and dendrites. Figure 2a,b illustrates a workflow for such connectome reconstruction. In this workflow, axons and dendritic shafts are reconstructed first (including branch points). Then a synapse movie mode is activated in webKnossos; in this mode, the user can fly along the pre-traced axon and click into the postsynaptic process whenever a synapse is encountered. In the final step, the postsynaptic partner (in about 90% a spine head, Supplementary Fig. 2a) is seeded for annotators to trace back to the main shaft of dendrites. In this workflow, the fraction of time spent on synapse annotation is small for sparse reconstructions (7–11% for typical network reconstructions, Fig. 2c, Supplementary Fig. 2b) but is becoming more substantial for dense reconstructions, approaching about 50% of reconstruction time (Supplementary Fig. 2b). Since synapse detection requires only a local image classification (unlike neurite tracing), automated synapse detection is likely to soon replace manual synapse detection in dense connectome reconstructions (e.g., refs. 16–19). To exemplify a full connectomic reconstruction using webKnossos, we finally reconstructed 497 dendrites (total path length of 93.6 mm, tracing redundancy 3), and determined all synapses with 32 axons from the training set (Fig. 1j, 4.55 mm path length of axons, tracing redundancy 6). We detected 104 synapses in this local connectome (Fig. 2d,e) (total annotation time was 27.3 h for axons, 133 h for dendrites, and 19.2 h for synapses).

In summary, webKnossos accelerates human 3D data interaction for EM-based connectomics in browser by about 4- to 13-fold, which likely saturates human interaction speed with 3D EM data of nervous tissue using flight mode. While tested on well-aligned 3D SBEM data from mammalian cortex, these results are expected to be comparable for other neuropil with comparable neurite morphology (especially branch-point rates, e.g., in subcortical structures and ganglion cells in mammalian retina, see <http://www.neuromorpho.org>). Reconstructions in highly anisotropic and potentially less well-aligned image data can still profit from the speedup because of faster display rates in ortho mode tracings (see Fig. 1i). Thus webKnossos can serve as a versatile high-efficiency tool for 3D image data annotation in various 3D image analysis settings in connectomics and other fields.

## METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

*Note: Any Supplementary Information and Source Data files are available in the online version of the paper.*

## ACKNOWLEDGMENTS

We thank D.Y. Buckley, E. Klinger and P. Laserstein for helpful discussions and comments on the manuscript; C. Roome, F. Kaiser and C. Guggenberger for outstanding IT support; and J. Striebel, S. Mischkewitz, N. Ring, A. Motta, B. Staffler and M. Zauser for contributions to the code. We thank E. Arzt and A.M. Burgin for hosting KMB at the Instituto de Investigación en Biomedicina de Buenos Aires, Argentina. We thank J. Abramovich, D. Acay, M. Ali, S. Babl, A. Banschbach, N. Berghaus, D. Beyer, L. Bezenberger, F. Bock, D. Boehm, A. Budco, L. Buxmann, D. Celik, H. Charif, N. Cipta, T. Decker, J. Depnering, V. Dienst, S. Dittmer, L. Dufter, J.E. Martinez, H. Feddersen, A. Feiler, L. Feist, L. Frey, K. Friedl, A. Gaebelein, L. Geiser, D. Greco, M. Groothuis, S. Gross, A. Haesl, S. Hain, J. Hartel, M.-L. Harwardt, B. Heftrich, M. Helbig, J. Heller, S. Hermann, J. Hoeltke, T. Hoermann, M. Hofweber, R. Hulse, R. Jakob, R. Jakoby, R. Janssen, M. Karabel, S. Kempf, L. Kirchner, J. Knauer, R. Kneissl, T. Koecke, P. Koenig, M. Kolodziej, F. Kraemer, E. Kuendiger, D. Kurt, M. Kuschneret, E. Laube, D. Laubender, M. Lehnardt, Irene Meindl, Iris Meindl, J. Meyer, P. Mueller, N. Neupaertl, M. Perschke, M. Poltermann, M. Praeve, K. Puechler, M. Reiner, T. Reimann, V. Robl, F. Ruecker, T. Ruff, F. Sahin, D. Schelii, V. Schuhbeck, S. Seiler, I. Stasiuk, B. Stiehl, A. Strubel, R. Thieleking, K. Trares, P. Wagner, Y. Wang, J. Wastl, A. Weber, S.S. Wehrheim, C. Weiss, M. Werr, A. Weyh, and J. Wiederspohn for annotation work.

## AUTHOR CONTRIBUTIONS

M.H. initiated and supervised the project; K.M.B., M.B., T.B., N.R., T.W. and M.H. developed specifications and conceptual design with contributions by H.W.; T.B., D.B., J.F., T.H., P.O., N.R., T.W., D.W., G.W. and K.M.B. implemented the software; H.W., M.B., K.M.B. and F.D. provided data; K.M.B., M.H., H.W. and M.B. analyzed the data; M.H., K.M.B. and M.B. wrote the manuscript with contributions by all authors.

## COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

- Denk, W. & Horstmann, H. *PLoS Biol.* **2**, e329 (2004).
- Eberle, A.L. *et al. J. Microsc.* **259**, 114–120 (2015).
- Kasthuri, N. *et al. Cell* **162**, 648–661 (2015).
- Helmstaedter, M. *Nat. Methods* **10**, 501–507 (2013).
- Saalfeld, S., Cardona, A., Hartenstein, V. & Tomancak, P. *Bioinformatics* **25**, 1984–1986 (2009).
- Schneider-Mizell, C.M. *et al. eLife* **5**, e12059 (2016).
- Rasmussen, L. *Lect. Notes Comput. Sci.* **3579**, 7 (2005).
- Helmstaedter, M., Briggman, K.L. & Denk, W. *Nat. Neurosci.* **14**, 1081–1088 (2011).
- Wanner, A.A., Genoud, C. & Friedrich, R.W. *Sci. Data* **3**, 160100 (2016).
- Wanner, A.A., Genoud, C., Masudi, T., Siksou, L. & Friedrich, R.W. *Nat. Neurosci.* **19**, 816–825 (2016).
- Helmstaedter, M. *et al. Nature* **500**, 168–174 (2013).
- Berning, M., Boergens, K.M. & Helmstaedter, M. *Neuron* **87**, 1193–1206 (2015).
- Hua, Y., Laserstein, P. & Helmstaedter, M. *Nat. Commun.* **6**, 7923 (2015).
- Braitenberg, V. & Schüz, A. *Anatomy of the cortex: statistics and geometry* Vol. 18 (Springer Science & Business Media, 2013).
- Narayanan, R.T. *et al. Cereb. Cortex* **25**, 4450–4468 (2015).
- Becker, C., Ali, K., Knott, G. & Fua, P. *IEEE Trans. Med. Imaging* **32**, 1864–1877 (2013).
- Kreshuk, A., Koethe, U., Pax, E., Bock, D.D. & Hamprecht, F.A. *PLoS One* **9**, e87351 (2014).
- Staffler, B. *et al. Preprint at* <http://biorxiv.org/content/early/2017/01/22/099994> (2017).
- Dorkenwald, S. *et al. Nat. Methods* **14**, 435–442 (2017).

## ONLINE METHODS

**3D SBEM data, animal experiments.** The 3D EM image data was acquired using serial blockface electron microscopy (SBEM<sup>1</sup>) from primary somatosensory cortex layer 4 of a P28 C57BL/6 male mouse, same data set as in ref. 12 ( $93 \times 60 \times 93 \mu\text{m}^3$ ,  $11.24 \times 11.24 \times 28 \text{ nm}^3$  resolution, data set 2012-09-28\_ex145\_07x2\_new2). All animal experiments were carried out with approval of the local animal research authorities (Regierungspräsidium Oberbayern, Germany) and in accordance with the German Animal Welfare Act.

**Image data transmission.** For efficient volume data transmission, data is requested in small cubic packages  $32^3$  voxel in size ('buckets') stored at original 8-bit depth. For bandwidth-limited settings, each voxel in a bucket is trimmed to the 4 most significant bits for transmission ('4-bit mode', user activatable). Buckets are requested along a priority ranking based on the current view point and direction of movement. In ortho mode, all buckets in the plane of the active viewport are loaded first, prioritized by the Manhattan distance to the viewport center. The buckets of the next two bucket layers in direction of movement are prioritized next, with two- and four-fold-reduced priority, respectively. All buckets are loaded at the user-specified magnification level.

In flight mode the preview volume is a square frustum with basis sized  $5 \times 5$  buckets, height of 2.5 buckets, oriented along the current movement direction, and top sized  $4 \times 4$  buckets. All buckets fully or partially contained in this volume are requested at original magnification, prioritized by the Manhattan distance between the respective bucket center and the current viewpoint. The prioritized bucket request queue is updated on each user movement.

### Measurement of 3D data transmission speed: webKnossos.

The speed of data transmission for 3D navigation in webKnossos (Fig. 1d) was measured as follows. webKnossos was run at a Hetzner (Gunzenhausen, Germany) data center on a server with the following specifications: Intel(R) Xeon(R) CPU E3-1245 V2 ( $4 \times 3.4 \text{ GHz}$ ); 32 GB RAM;  $15 \times 3 \text{ TB}$  HWRAID HDD. The EM data set was viewed in 4-bit mode in webKnossos run in Google Chrome (version 56) on a computer in the MPI for Brain Research. A neurite was picked and followed using the forward and arrow keys, keeping the forward key pressed where possible (which resulted in continuous image stream in webKnossos). To emulate reduced connectivity settings, the developer tools function of Google Chrome was used. The 'transcontinental' experiment (Fig. 1c) was performed on a computer connected to the network of the Instituto de Investigación en Biomedicina de Buenos Aires, Argentina, accessing the webKnossos instance running on the Hetzner server in Germany (see above).

### Measurement of 3D data transmission speed: sequential 2D.

The speed of data transmission for sequential 2D image navigation (Fig. 1c) was measured as follows. We followed the instructions by the CATMAID authors to optimize server performance (published under [https://groups.google.com/forum/#!topic/catmaid/vE\\_4iLrPv4](https://groups.google.com/forum/#!topic/catmaid/vE_4iLrPv4)). A CATMAID instance (version 2016.12.16) was installed on a server in the compute center of the Max Planck Society (Garching, Germany) with the following specifications: Xeon E5-2630 12 cores, 128 GB RAM, 10 Gb network, JBOD of  $4 \times$  Intel DC S3500 240 GB SSDs, Ubuntu 14.04. Postgres and

data partitions resided on SSDs; XFS was used with noatime. The 3D image data set 2012-09-28\_ex145\_07x2\_new2 (s. above) was converted to a series of  $256 \times 256 \text{ px}$  jpg images, compressed by 75% with jpg headers removed to further reduce file size as suggested by the CATMAID authors. The data were resliced into three image series along the three cardinal directions.

The 'transcontinental' experiment was performed by accessing CATMAID on <https://www.openconnectomeproject.org> from a desktop computer in the MPI for Brain Research and by accessing the custom CATMAID instance at the Max Planck datacenter in Germany from a computer connected to the network of the Instituto de Investigación en Biomedicina de Buenos Aires, Argentina. All tests were performed in Google Chrome (version 56). The viewports in CATMAID and webKnossos were set to similar size.

**Annotator training.** For training annotators at high-speed annotation (Fig. 1h), 40 training neurites were selected from the cortex data set (see above). For this, a bounding box sized  $4.5 \times 4.5 \times 4.2 \mu\text{m}^3$  was chosen. Then, two annotators were asked to reconstruct all processes within this bounding box. Next, each of the reconstructed processes was classified as axon, dendrite or glia. Finally, 40 of the 68 processes classified as axons were randomly selected. For each process, an expert annotator defined a starting position and a starting direction (required for flight-mode annotation).

51 annotators were trained. These annotators were randomly assigned to two groups (flight ( $n = 26$ ) and ortho mode ( $n = 25$ )). Annotators were asked to watch an introductory video, which instructed them to increase their maximum velocity setting (the speed at which the annotator progresses through the data when the space key is held down continuously). Each annotator was presented with the 40 training processes in random order. The annotators' preset maximum velocities were monitored during the annotation process. If the ratio of tracing speed and preset maximum velocity was higher than 0.75 for an entire tracing, the annotator was notified via e-mail and asked to increase the maximum velocity setting for the next annotation.

**Axon test reconstruction.** For the test of axon reconstruction speed (Fig. 1i), the 51 previously trained annotators were asked to reconstruct 20 randomly selected axons 8 weeks after the initial training. 26 annotators signed up for this experiment (13 that had been trained on flight mode and 13 that had been trained on ortho mode). These annotators had not been faster in the final ten training iterations than the whole group of annotators ( $P = 0.246$  (Wilcoxon rank-sum test) and  $P = 0.250$  ( $t$ -test)) and had not been faster in the final training iteration ( $P = 0.699$  (Wilcoxon rank-sum test) and  $P = 0.649$   $t$ -test).

To select a set of representative axons, a  $(2.5 \mu\text{m})^3$  bounding box (located randomly within the cuboid of  $15 \mu\text{m}$  edge length centered to the data set center) was chosen that did not contain a soma, and all neuronal and glial processes within this bounding box were reconstructed by one expert annotator. Then all processes were classified as axonal, dendritic or glial. Three additional expert annotators proofread the annotation. Then, 20 of the 41 processes labeled as axonal were randomly selected, and for each axon a seed position and initial orientation were defined within the bounding box.

The 26 annotators were first asked to again reconstruct five neurites from the training experiment (these five neurites were randomly chosen from the 40 training seeds and were the same

for all annotators; the sequence in which these were presented was randomized for each annotator). Then each annotator was asked to reconstruct the 20 previously unseen test axons (in a sequence randomized per annotator) in flight mode or ortho mode. After all annotators had finished, all annotations were automatically scanned for open branch points (i.e., positions at which the annotator had set a branch point flag but had forgotten to jump back to for inspection) and seed nodes with a degree of 1 (i.e., starting points which had only been traced in one direction). 17 open branch points (12 at first node) and 22 unidirectional seeds were detected (5 in flight, 17 in ortho) of 2,170 fully annotated branch points, total. In these cases the annotators were asked to go back to the task and continue their annotations. The code for this automated annotation checking is provided in **Supplementary Software 1**.

**Dendrite reconstruction.** 497 dendrites were reconstructed in flight mode by 47 of the annotators previously trained in ortho or flight mode (see above). Those annotators that had worked in ortho mode before were asked to watch the instruction movie for flight mode before performing the dendrite reconstructions. Annotators were instructed to set the data set quality setting in webKnossos to medium (which means that image data is displayed at lower resolution) and not to reconstruct spines.

The 497 dendrite seeds for the connectome reconstruction (Fig. 2) were drawn randomly from a set of over 2,000 dendrites that had previously been reconstructed using webKnossos. For all dendrites, the z-axis pointing toward the data set center was used as initial flight orientation.

**Measurement of annotation speed.** For measurement of annotation speed, the path length of a given neurite and the time it took to annotate that neurite were determined. To measure neurite path length from a skeleton annotation, the lengths of all edges within a skeleton were summed (as in refs. 8 and 11). However, this method has two caveats. First, noise in the placement of skeleton nodes will be biased to only increase apparent skeleton length, not decrease it, which could potentially lead to an overestimation of annotation speed. Second, this effect will depend on the density of placed skeleton nodes. Since in flight mode the skeleton nodes are placed automatically, the density of skeleton nodes is substantially higher in flight mode than in ortho mode tracings (flight,  $6.3 \pm 0.68$  nodes/ $\mu\text{m}$ ; ortho,  $1.79 \pm 0.67$  nodes/ $\mu\text{m}$ , measured on the ten axons used for Fig. 11-o). To account for these potential biases, we first used nonuniform rational b-spline (NURBS<sup>20</sup>)-based skeleton smoothing to calibrate the effect of node placement noise on skeleton path length (**Supplementary Fig. 1f**); using the skeleton nodes as support knots, NURBS spline order (i.e., the degree of smoothing)  $NO = 4$  and clamping the first and last node. We found that post-NURBS path length measurements of flight tracings are still on average  $14.93 \pm 1.08\%$  (mean  $\pm$  s.e.m.) longer than ortho tracings (**Supplementary Fig. 1g**). To correct for this and for the difference in node densities between tracing modes, we scaled  $NO$  in dependence of skeleton node density  $D_s$  (in number of nodes per  $\mu\text{m}$  edge-based skeleton length),

$$NO = \min \left( \text{ceil} \left( c_1 \left( \frac{D_s}{c_2} \right)^{c_3} \right), N_n \right)$$

with  $N_n$  (the total number of skeleton nodes per tracing) and parameters  $c_1$ ,  $c_2$  and  $c_3$ . The parameters were adjusted such that the average path lengths of neurites from the training set were similar for flight- and ortho-mode tracings (resulting in  $c_1 = 50$ ,  $c_2 = 5$ ,  $c_3 = 4$ ; as a result, average ortho and flight path length for a given axon agreed within  $1.79 \pm 1.16\%$ ). See **Supplementary Figure 1g** and **Supplementary Software 1–3** for comparisons of skeleton path length measurements based on edge length and NURBS smoothing with fixed  $NO$  and variable  $NO$ , respectively. The length measurements involving NURBS smoothing with fixed  $NO$  and variable  $NO$  reduced the path length obtained from the simple edge length addition method by less than 20% (**Supplementary Fig. 1g**). The variable  $NO$  path length measurement method was used for speed measurements in the axon test set and the dendrite tracings.

To determine the annotation time of a given tracing, the administrative API of webKnossos was used (**Supplementary Software 4**) to log autosave events. Autosave events are triggered when the annotator is actively tracing within the last 30 s and the last autosave was more than 30 s ago. Therefore, during annotation work, an autosave is submitted every 30 s (but not during pauses the annotator chooses to take). Annotation time was measured as the number of autosave events times 30 s. This is also the time used for determining annotator payment.

**Annotation redundancy: RESCOP.** For determining the dependence of annotation error rates on annotation redundancy (**Fig. 11-o**), multiple annotations of the same neurite from different annotators were consolidated using RESCOP<sup>8</sup>. Briefly, the priors and decision boundaries were fitted separately for axons traced in ortho and flight mode (**Supplementary Fig. 1h,i**). The priors were fitted using 20 randomly selected annotations for each neurite from the training annotations (i.e., 800 annotations for ortho and flight mode, respectively, total of 865,121 edges, 866,721 nodes for computing the vote histogram, **Supplementary Fig. 1h**). The resulting decision boundaries are shown in **Supplementary Figure 1i**.

**Measurement of annotation error rates.** For the measurement of annotation errors, 10 of the 20 test axons (**Fig. 1j**) were randomly selected. For these ten axons, a ground truth annotation was generated. To do this, the axon was first traced by one expert annotator. Then, all annotations of this axon from all tracers and tracing modes were superimposed; and all locations of discrepancy between the experts' annotation and all other annotations were inspected. Remaining errors in the expert annotation were corrected. Finally, two additional experts verified the ground truth annotation independently.

Then, for each of the ten axons, three ortho-mode and three flight-mode annotations were randomly selected and their discrepancies to the ground truth annotation counted as in ref. 8 (Fig. 5c in this reference). Similarly, consensus skeletons at redundancies 2, 3, 4, 5, 6, 7, 10 and 13 were computed using RESCOP<sup>8</sup> (see above) for each of the ten axons and the two tracing modes, respectively. For each redundancy, 3 sets of tracings were randomly drawn from the available 13 tracings per axon and tracing mode. Thus, together, 540 reconstructions were error analyzed.

Error analysis was done by one expert annotator and proof-read by a second expert annotator. Both experts were blinded to the tracing mode in which the reconstructions were performed.



For error analysis, the reconstructions were plotted in the three cardinal projections, overlaid with the ground truth reconstruction. Errors were classified into missing branches (false negatives) and wrongly added branches (false positives). Jumps from one process into another were counted twice, once as FP and once as FN. Errors were further classified according to the length of added or omitted neurite pieces ( $>10\ \mu\text{m}$ ,  $5\text{--}10\ \mu\text{m}$ ,  $3\text{--}5\ \mu\text{m}$ ,  $1\text{--}3\ \mu\text{m}$ ; discrepancies smaller than  $1\ \mu\text{m}$  were not counted, as in ref. 8). Error segments larger than  $10\ \mu\text{m}$  were classified as continuation errors (Fig. 1k).

For the measurement of errors in dendrite reconstructions, 10 of the 497 reconstructed dendrites were randomly selected and error-annotated as for the axon reconstructions. Since error rates were substantially lower for dendrites than for axons (Fig. 1l–o), only redundancies 1–6 were evaluated for dendrites.

**Path length of consolidated reconstructions.** To make our results comparable to error rates reported in refs. 8, 11 and 13, we normalized the number of errors to the path length of the ground truth skeleton for each axon (Fig. 1l,m). However, since some annotations were shorter (due to missed neurite pieces) and others longer (due to added neurite pieces), we wanted to assure that our conclusions about error rates (Fig. 1l–o) were still correct when instead the neurite path length of the actual tracing or consolidation was used for error rate computation. To do so, we determined the path length for each RESCOP-consolidated reconstruction by generating a version of the ground truth reconstruction that matched the respective RESCOP-consolidated reconstruction (including its possible false negative errors), and we measured that skeleton's path length as described above.

**Synapse annotation and connectome reconstruction.** To exemplify the full analysis workflow for reconstructing connectomes using webKnossos, we used all axons from the training reconstructions (32 axons, at RESCOP redundancy 6, step I in Fig. 2a–c) and 497 dendrite reconstructions (Fig. 2d, at redundancy 3, step II in Fig. 2a–c). For synapse annotation (step III in Fig. 2a–c), a synapse movie mode in webKnossos was used (this mode is automatically activated for webKnossos tasks of type 'synapseannotation'). This was built as an extension of flight mode in which the previously reconstructed skeleton was displayed. The annotator was asked to mark synapses by setting a single node into the postsynaptic process while navigating along the axon. For the synapse movie mode, the (consolidated) reconstruction was first cut into unbranched parts, and each of these parts was presented to the annotators (see **Supplementary Software 1** for the corresponding MATLAB code).

Ten annotators were trained for synapse annotation in an introductory 1-h seminar followed by two training axons for which they received immediate feedback. Then all annotators were asked to determine the output synapses of all 32 axons. To measure the precision and recall of synapse detection by student annotators, four of the axons were randomly selected, and synapse detection errors were determined by expert annotators. The student annotator with optimal precision and recall of synapse detection (precision 96%, recall 89%) was selected for the generation of the output connectome. In addition, annotators were instructed to mark axons as putative inhibitory axons if the majority of output synapses were made onto shafts.

For the axons that the best annotator marked as inhibitory, a second annotator was asked to annotate the synapses of that axon. For the annotation of inhibitory synapses, the annotator was instructed not to focus on speed of synapse annotation. The resulting synapse annotations were reviewed by an expert annotator to establish error rates for inhibitory synapse annotation (no error in 20 reviewed synapse annotations).

This procedure operated at  $1.2 \pm 0.5\ \text{h per mm}$  candidate axon segment length ( $n = 151$ , excitatory axons;  $1.8 \pm 1.0\ \text{h per mm}$  for all axons,  $n = 178$ ).

To determine whether the postsynaptic targets of the reconstructed axons matched any of the 497 dendrites in the connectome, the annotation of the postsynaptic partner in synapse mode was used as a new seed for an annotation task (step IV in Fig. 2a–c). The annotators for these tasks were asked to only reconstruct the postsynaptic structure (in about 90% of cases a spine) in ortho mode until it entered a dendritic shaft and to then place three additional skeleton nodes in the shaft center to simplify the matching to dendrite reconstructions. This annotation had a consumption of  $31.1 \pm 28.0\ \text{s}$  annotation time per spine, mean  $\pm$  s.d.,  $n = 975$ ; i.e.  $2.3 \pm 1.3\ \text{h per mm}$  axon path length. Error rates of this postsynaptic process annotation were established by inspection of 30 randomly selected postsynaptic structures by an expert annotator (one wrong annotation).

To match the postsynaptic partner reconstructions (step V in Fig. 2a) with the 497 dendrite reconstructions, we finally measured the average distance  $d_{pd}$  between all dendrites and the three shaft nodes of each postsynaptic partner reconstruction, and we detected the dendrite with the smallest average distance. To determine an attachment threshold—i.e. a maximum average distance  $d_{pd^*}$  up to which a postsynaptic partner reconstruction was considered to match a dendrite reconstruction—we used a randomly chosen set of 200 partner reconstructions. In these, the distribution of  $d_{pd}$  (**Supplementary Fig. 2a**) indicated a threshold distance  $d_{pd^*}$  of 250 nm. To determine the error rate of postsynaptic partner matching, we evaluated the matching in an additional set of 200 randomly chosen spines and their closest dendrite (21 true positives, 1 false positive, 178 true negatives, no false negatives). All code for these procedures is available in **Supplementary Software 1**.

**Connectome annotation time estimates.** For the annotation time approximation of an example L2/3–L4 cortical connectome (**Supplementary Fig. 2c**), we used  $1.5\ \text{mm/h}$  reconstruction speed and six-fold redundancy for axons, and  $2.1\ \text{mm/h}$  reconstruction speed at three-fold redundancy for dendrites. For estimating the reconstruction time spent on synapse annotation (**Supplementary Fig. 2b**), two approaches for synapse annotation were considered. One, axon-based synapse annotation (Fig. 2a,b), proceeds along axons, marking synapses and identifying postsynaptic partners, which are then matched to dendrite reconstructions (see “Results”). The other, dendrite-based synapse annotation, proceeds along dendrites, reconstructing all spines along dendritic shafts. Spine annotation along dendrites proceeds at about 40 s per spine (time taken to reconstruct a spine and mark its presynaptic partner) at a spine density of about 1 per  $\mu\text{m}$  dendrite length. In both strategies, we assumed that only proximities of axons and dendrites at less than  $5\ \mu\text{m}$  distance need to be investigated for synapses. Therefore, depending on the density of axons and

dendrites in a given reconstruction task, the length of axons and dendrites that need to be synapse searched varies, which yields an optimal strategy for any given volume density of axons and dendrites. Dendrite and axon reconstruction speed used for these estimates was 2.1 mm/h and 1.5 mm/h (at three- and six-fold redundancy, respectively).

**Statistical tests.** The speed and the error comparison between ortho and flight tracers and the speed comparison between the subset of annotators for the second experiment and all annotators used Wilcoxon rank-sum test. The speed comparison for annotators between beginning and end of training used Wilcoxon signed-rank test. The correlation between tracing error rate and tracing speed was computed using Pearson's correlation.

**Software availability, code availability and licensing.** webKnossos is available for testing at <https://demo.webknossos.brain.mpg.de> together with example data sets: the published retina data sets e2198 (ref. 21), k0563 (refs. 8, 11 and 21) and e2006 (ref. 11), a  $20 \times 20 \times 20 \mu\text{m}^3$  sized subvolume of the data set 2012-09-28\_ex145\_07x2\_new2 used for webKnossos testing (see above), and an example fluorescence data set (FD0149-2, data not shown). See **Supplementary Video 2** for an introductory video.

The webKnossos source code is provided as **Supplementary Software 5** and is also available at <https://github.com/scalableminds/webKnossos>. webKnossos is licensed under the AGPLv3 license (this applies to all source code files in **Supplementary Software 1–5** and GitHub repository). webKnossos uses the following software packages and technologies: Scala, JDK 8, Play, mongoDB, WebGL, ThreeJS, Backbone, sbt.

**Data availability statement.** webKnossos is openly accessible at <https://demo.webknossos.brain.mpg.de>, where data sets from retina and cortex can be browsed and annotated. The entire SBEM data set of the mouse cortex that support the findings in this study are available from the corresponding author upon reasonable request. webKnossos is open source, source code is available as **Supplementary Software 5** and at <https://github.com/scalableminds/webknossos>. All reconstructions used in this study are available in **Supplementary Software 1, 2 and 3**. Source data for **Figures 1 and 2** are available online.

20. Piegl, L. & Tiller, W. *The NURBS book* (Springer Science & Business Media, 2012).
21. Briggman, K.L., Helmstaedter, M. & Denk, W. *Nature* **471**, 183–188 (2011).

## 6.3 SynEM: Automated synapse detection for connectomics

**Abstract** Nerve tissue contains a high density of chemical synapses, about 1 per  $\mu m^3$  in the mammalian cerebral cortex. Thus, even for small blocks of nerve tissue, dense connectomic mapping requires the identification of millions to billions of synapses. While the focus of connectomic data analysis has been on neurite reconstruction, synapse detection becomes limiting when datasets grow in size and dense mapping is required. Here, we report SynEM, a method for automated detection of synapses from conventionally en-bloc stained 3D electron microscopy image stacks. The approach is based on a segmentation of the image data and focuses on classifying borders between neuronal processes as synaptic or non-synaptic. SynEM yields 97% precision and recall in binary cortical connectomes with no user interaction. It scales to large volumes of cortical neuropil, plausibly even whole-brain datasets. SynEM removes the burden of manual synapse annotation for large densely mapped connectomes.

**Contributions** Conceived and initiated the project: MH; supervised the project: MH and PvdS; Developed algorithms, implemented algorithms, analyzed data: BS; provided segmentations and contributed to algorithm development: MB; provided EM data: KMB; provided expert synapse annotations: AG; wrote the paper: MH and BS.

**Copyright** This article [Staffler et al., 2017] is licensed under the CC BY 4.0.

**Note** Accepted manuscript, PDF only. Full online edition to follow.

# 1 SynEM, Automated synapse 2 detection for connectomics

---

3 Benedikt Staffler<sup>1</sup>, Manuel Berning<sup>1</sup>, Kevin M. Boergens<sup>1</sup>, Anjali Gour<sup>1</sup>, Patrick van  
4 der Smagt<sup>2</sup>, Moritz Helmstaedter<sup>1</sup>

5 <sup>1</sup>Department of Connectomics, Max Planck Institute for Brain Research, D-60438  
6 Frankfurt, Germany

7 <sup>2</sup>Biomimetic Robotics and Machine Learning, TUM, D-80333 Munich, Germany;  
8 current affiliation: Data Lab, VW Group, 80805 Munich, Germany

9

10

11

12

13

14

15 Correspondence: B.S. ([benedikt.staffler@brain.mpg.de](mailto:benedikt.staffler@brain.mpg.de)) or M.H. ([mh@brain.mpg.de](mailto:mh@brain.mpg.de))

16

17

18

19

## ABSTRACT

Nerve tissue contains a high density of chemical synapses, about 1 per  $\mu\text{m}^3$  in the mammalian cerebral cortex. Thus, even for small blocks of nerve tissue, dense connectomic mapping requires the identification of millions to billions of synapses. While the focus of connectomic data analysis has been on neurite reconstruction, synapse detection becomes limiting when datasets grow in size and dense mapping is required. Here, we report SynEM, a method for automated detection of synapses from conventionally en-bloc stained 3D electron microscopy image stacks. The approach is based on a segmentation of the image data and focuses on classifying borders between neuronal processes as synaptic or non-synaptic. SynEM yields 97% precision and recall in binary cortical connectomes with no user interaction. It scales to large volumes of cortical neuropil, plausibly even whole-brain datasets. SynEM removes the burden of manual synapse annotation for large densely mapped connectomes.

## 38 INTRODUCTION

39

40 The ambition to map neuronal circuits in their entirety has spurred substantial  
 41 methodological developments in large-scale 3-dimensional microscopy (Denk &  
 42 Horstmann, 2004, Hayworth et al., 2006, Knott et al., 2008, Eberle et al., 2015),  
 43 making the acquisition of datasets as large as 1 cubic millimeter of brain tissue or  
 44 even entire brains of small animals at least plausible (Mikula et al., 2012, Mikula &  
 45 Denk, 2015). Data analysis, however, is still lagging far behind (Helmstaedter, 2013).  
 46 One cubic millimeter of gray matter in the mouse cerebral cortex, spanning the entire  
 47 depth of the gray matter and comprising several presumed cortical columns (Fig. 1a),  
 48 for example, contains at least 4 kilometers of axons, about 1 kilometer of dendritic  
 49 shafts, about 1 billion spines (contributing an additional 2-3 kilometers of spine neck  
 50 path length) and about 1 billion synapses (Fig. 1b). Initially, neurite reconstruction  
 51 was so slow, that synapse annotation comparably paled as a challenge (Fig. 1c):  
 52 when comparing the contouring of neurites (proceeding at 200-400 work hours per  
 53 millimeter neurite path length) with synapse annotation by manually searching the  
 54 volumetric data for synaptic junctions (Fig. 1d, proceeding at about 0.1 hour per  
 55  $\mu\text{m}^3$ ), synapse annotation consumed at least 20-fold less annotation time than  
 56 neurite reconstruction (Fig. 1c). An alternative strategy for manual synapse detection  
 57 is to follow reconstructed axons (Fig. 1e) and annotate sites of vesicle accumulation  
 58 and postsynaptic partners. This axon-focused synapse annotation reduces synapse  
 59 annotation time by about 8-fold for dense reconstructions (proceeding at about 1 min  
 60 per potential contact indicated by a vesicle accumulation, which occurs every about  
 61 4-10  $\mu\text{m}$  along axons in mouse cortex).

62 With the development of substantially faster annotation strategies for neurite  
 63 reconstruction, however, the relative contribution of synapse annotation time to the  
 64 total reconstruction time has substantially changed. Skeleton reconstruction  
 65 (Helmstaedter et al., 2011) together with automated volume segmentations  
 66 (Helmstaedter et al., 2013, Berning et al., 2015), allow to proceed at about 7-10  
 67 hours per mm path length (mouse retina, Helmstaedter et al., 2013) or 4-7 hours per  
 68 mm (mouse cortex, Berning et al., 2015), thus about 50-fold faster than manual  
 69 contouring. Recent improvements in online data delivery and visualization (Boergens  
 70 et al., 2017) further reduce this by about 5-10 fold. Thus, synapse detection has  
 71 become a limiting step in dense large-scale connectomics. Importantly, any further

improvements in neurite reconstruction efficiency would be bounded by the time it takes to annotate synapses. Therefore, automated synapse detection for large-scale 3D EM data is critical.

High-resolution EM micrographs are the gold standard for synapse detection (Gray, 1959, Colonnier, 1968). Images acquired at about 2-4 nm in-plane resolution have been used to confirm chemical synapses using the characteristic intense heavy metal staining at the postsynaptic membrane, thought to be caused by the accumulated postsynaptic proteins (“postsynaptic density”, PSD), and an agglomeration of synaptic vesicles at the membrane of the presynaptic terminal. While synapses can be unequivocally identified in 2-dimensional images when cut perpendicularly to the synaptic cleft (Fig. 1f), synapses at oblique orientations or with a synaptic cleft in-plane to the EM imaging are hard or impossible to identify. Therefore, the usage of 3D EM imaging with a high resolution of 4-8 nm also in the cutting dimension (FIB/SEM, Knott et al., 2008) is ideal for synapse detection. For such data, automated synapse detection is available and successful (Kreshuk et al., 2011, Becker et al., 2012, 2013, Suppl. File 1). However, FIB-SEM currently does not scale to large volumes required for connectomics of the mammalian cerebral cortex. Serial Blockface EM (SBEM, Denk & Horstmann, 2004) scales to such mm<sup>3</sup>-sized volumes. However, SBEM provides a resolution just sufficient to follow all axons in dense neuropil and to identify synapses across multiple sequential images, independent of synapse orientation (Fig. 1g, see also Synapse Gallery in Supplementary File 4; the resolution of SBEM is typically about 10x10x30 nm<sup>3</sup>; Fig. 1g). In this setting, synapse detection methods developed for high-in plane resolution data do not provide the accuracy required for fully automated synapse detection (see below).

Here we report SynEM, an automated synapse detection method based on an automated segmentation of large-scale 3D EM data (using SegEM, Berning et al., 2015; an earlier version of SynEM was deposited on biorxiv, Staffler et al., 2017). SynEM is aimed at providing fully automated connectomes from large-scale EM data in which manual annotation or proof reading of synapses is not feasible. SynEM achieves precision and recall for single-synapse detection of 88% and for binary neuron-to-neuron connectomes of 97% without any human interaction, essentially removing the synapse annotation challenge for large-scale mammalian connectomes.

## RESULTS

### Interface classification

We consider synapse detection as a classification of interfaces between neuronal processes as synaptic or non-synaptic (Fig. 2a; see also Mishchenko et al., 2010, Kreshuk et al., 2015, Huang et al., 2016). This approach relies on a volume segmentation of the neuropil sufficient to provide locally continuous neurite pieces (such as provided by SegEM, Berning et al., 2015, for SBEM data of mammalian cortex), for which the contact interfaces can be evaluated.

The unique features of synapses are distributed asymmetrically around the synaptic interface: presynaptically, large vesicle pools extend into the presynaptic terminal over at least 100-200 nm; postsynaptically, the PSD has a width of about 20-30 nm. To account for this surround information our classifier considers the subvolumes adjacent to the neurite interface explicitly and separately, unlike previous approaches (Kreshuk et al., 2015, Huang et al., 2016), up to distances of 40, 80, and 160 nm from the interface, restricted to the two segments in question (Fig. 2b; the interface itself was considered as an additional subvolume). We then compute a set of 11 texture features (Table 1, this includes the raw data as one feature), and derive 9 simple aggregate statistics over the texture features within the 7 subvolumes. In addition to previously used texture features (Kreshuk et al., 2011, Table 1), we use the local standard deviation, an intensity-variance filter and local entropy to account for the low-variance (“empty”) postsynaptic spine volume and presynaptic vesicle clouds, respectively (see Fig. 2c for filter output examples and Fig. 2d for filter distributions at an example synaptic and non-synaptic interface). The “sphere average” feature was intended to provide information about mitochondria, which often impose as false positive synaptic interfaces when adjacent to a plasma membrane. Furthermore, we employ 5 shape features calculated for the border subvolume and the two subvolumes extending 160 nm into the pre- and postsynaptic processes, respectively. Together, the feature vector for classification had 3224 entries for each interface (Table 1).



## SynEM workflow and training data

We developed and tested SynEM on a dataset from layer 4 (L4) of mouse primary somatosensory cortex (S1) acquired using SBEM (dataset ex145\_07x2, Boergens et al., in prep.; the dataset was also used in developing SegEM, Berning et al., 2015). The dataset had a size of  $93 \times 60 \times 93 \mu\text{m}^3$  imaged at a voxel size of  $11.24 \times 11.24 \times 28 \text{ nm}^3$ . The dataset was first volume segmented (SegEM, Berning et al., 2015, Fig. 2a, see Fig. 2e for a SynEM workflow diagram). Then, all interfaces between all pairs of volume segments were determined, and the respective subvolumes were defined. Next, the texture features were computed on the entire dataset and aggregated as described above. Finally, the shape features were computed. Then, the SynEM classifier was implemented to output a synapse score for each interface and each of the two possible pre-to-postsynaptic directions (Fig. 3a-c). The SynEM score was then thresholded to obtain an automated classification of interfaces into synaptic / non-synaptic ( $\theta$  in Fig. 3a). Since the SynEM scores for the two possible synaptic directions at a given neurite-to-neurite interface were rather disjunct in the range of relevant thresholds, we used the larger of the two scores for classification (Fig. 3b;  $\theta_s$  and  $\theta_{nn}$  refer to the SynEM thresholds optimized for single synapse or neuron-to-neuron connectome reconstruction, respectively, see below).

We obtained labels for SynEM training and validation by presenting raw data volumes of  $(1.6 \times 1.6 \times 0.7-1.7) \mu\text{m}^3$  that surrounded the segment interfaces to trained student annotators (using a custom-made annotation interface in Matlab, Fig. 3 – figure supplement 1). The raw data was rotated such that the interface was most vertically oriented in the image plane presented to the annotators; the two interfacing neurite segments were colored transparently for identification (this could be switched off by the annotators when inspecting the synapse, see Methods for details). Annotators were asked to categorize the presented interface as either non-synaptic, pre-to-postsynaptic, or post-to-presynaptic (Fig. 3c, Fig. 3 – figure supplement 1). The synaptic labels were then verified by an expert neuroscientist. A total of 75,383 interfaces (1,858 synaptic, 73,525 non-synaptic) were annotated in image volumes drawn from 40 locations within the entire EM dataset (Fig. 3 – figure supplement 2). About 80% of the labels (1467 synaptic, 61,619 non-synaptic) were used for training, the remaining were used for validation.

Initially, we interpreted the annotator's labels in an undirected fashion: irrespective of synapse direction, the label was interpreted as synaptic (and non-synaptic otherwise, Fig. 3c, "Undir."). We then augmented the training data by including mirror-reflected copies of the originally presented synapses, maintaining the labels as synaptic (irrespective of synapse direction) and non-synaptic (Fig. 3c, "Augmented"). Finally, we changed the labels of the augmented training data to reflect the direction of synaptic contact: only synapses in one direction were labeled as synaptic, and non-synaptic in the inverse direction (Fig. 3c "Directed").

## **SynEM evaluation**

Fig. 3d shows the effect of the choice of features, aggregate statistics, classifier parameters and label types on SynEM precision and recall. Our initial classifier used the texture features from Kreshuk et al., 2011 with minor modifications and in addition the number of voxels of the interface and the two interfacing neurite segmentation objects (restricted to 160 nm distance from the interface) as a first shape feature (Table 1). This classifier provided only about 70% precision and recall (Fig. 3d). We then extended the feature space by adding more texture features capturing local image statistics (Table 1) and shape features. In particular, we added filters capturing local image variance in an attempt to represent the "empty" appearance of postsynaptic spines, and the presynaptic vesicle clouds imposing as high-frequency high-variance features in the EM images. Also, we added more subvolumes over which features were aggregated (see Fig. 2b), increasing the dimension of the feature space from 603 to 3224. Together with additional aggregate statistics, the classifier reached about 75% precision and recall. A substantial improvement was obtained by switching from an ensemble of decision-stumps (one-level decision tree) trained by AdaBoostM1 (Freund & Schapire, 1997) as classifier to decision stumps trained by LogitBoost (Friedman et al., 2000). In addition, the directed label set proved to be superior. Together, these improvements yielded a precision and recall of 87% and 86% on the validation set (Fig. 3d).

We then evaluated the best classifier from the validation set (Fig. 3d, 'Direct & Logit') on a separate test set. This test set was a dense volume annotation of all synapses in a randomly positioned region containing dense neuropil of size  $5.8 \times 5.8 \times 7.2 \mu\text{m}^3$  from the L4 mouse cortex dataset. All synapses were identified by 2 experts, which included the reconstruction of all local axons, and validated once more by another expert on a subset of synapses. In total, the test set contained 235 synapses and 20319 non-synaptic interfaces. SynEM automatically classified these at 88%

precision and recall (Fig. 3e, F1 score of 0.883). Since the majority of synapses in the cortex are made onto spines we also evaluated SynEM on all spine synapses in the test set (n=204 of 235 synapses, 87%, Fig. 3e). On these, SynEM performed even better, yielding 94% precision and 89% recall. (Fig. 3e, F1 score of 0.914 ).

## **Comparison to previous methods**

We next compared SynEM to previously published synapse detection methods (Fig. 3f, Mishchenko et al., 2010, Kreshuk et al., 2011, Kreshuk et al., 2014, Becker et al., 2012, Roncal et al., 2015, Dorkenwald et al., 2017). Other published methods were either already shown to be inferior to one of these approaches (Perez et al., 2014, Marquez Neila et al., 2016) or developed for specific subtypes of synapses, only (Jagadeesh et al., 2014, Plaza et al., 2014, Huang et al., 2016); these were therefore not included in the comparison. SynEM outperforms the state-of-the-art methods when applied to our SBEM data acquired at 3537 nm<sup>3</sup> voxel size (Fig. 3f, Fig. 3 – figure supplement 3). In addition, we applied SynEM to a published 3D EM dataset acquired at more than 10-fold smaller voxel size (3 x 3 x 30 = 270 nm<sup>3</sup>) using automated tape-collecting ultramicrotome-SEM imaging (ATUM, Kasthuri et al., 2015). SynEM also outperforms the method developed for this data (VesicleCNN, Roncal et al., 2015; Fig. 3f and Fig. 3 – figure supplement 4), indicating that SynEM is applicable to EM data of various modalities and resolution.

It should furthermore be noted that for connectomics, in addition to the detection of the location of a synapse, the two neuronal partners that form the synapse and the direction of the synapse have to be determined. The performance of the published methods as reported in Fig. 3f only include the synapse detection step. Interestingly, the recently published method (Dorkenwald et al., 2017) reported that the additional detection of the synaptic partners yielded a drop of performance of 3% precision and 10% recall (F1 score decreased by about 5% from 0.906 to 0.849) compared to synapse detection alone (Fig. 3f, see Dorkenwald et al., 2017). This indicates that the actual performance of this method on our data would be lower when including partner detection. SynEM, because of the explicit classification of directed neurite interfaces, in contrast, explicitly provides synapse detection, partner detection and synapse directionality in one classification step.

## Remaining SynEM errors, feature importance, and computational feasibility

Fig. 4a shows examples of correct and incorrect SynEM classification results (evaluated at  $\theta_s$ ). Typical sources of errors are vesicle clouds close to membranes that target nearby neurites (Fig. 4a, FP), Mitochondria in the pre- and/or postsynaptic process, very small vesicle clouds and/or small PSDs (Fig. 4a, FN), and remaining SegEM segmentation errors. To estimate the effect of segmentation errors on SynEM performance, we investigated all false positive and false negative detections in the test set and checked for the local volume segmentation quality. We found that, in fact, 26 of the 28 FNs and 22 of the 27 FPs were at locations with a SegEM error in proximity. Correcting these errors also corrected the SynEM errors in 22 of 48 (46%) of the cases. This indicates that further improvement of volume segmentation can yield an even further reduction of the remaining errors in SynEM-based automated synapse detection.

We then asked which of the SynEM features had highest classification power, and whether the newly introduced texture and shape features contributed to classification. Boosted decision-stump classifiers allow the ranking of features according to their classification importance (Fig. 4b). 378 out of 3224 features contributed to classification (leaving out the remaining features did not reduce accuracy). The 10 features with highest discriminative power (Table 2) in fact contained two of the added texture filters (int-var and local entropy) and a shape feature. The three most distinctive subvolumes (Fig. 4b) were the large presynaptic subvolume, the border and the small postsynaptic subvolume. This suggests that the asymmetry in pre- vs. postsynaptic aggregation volumes in fact contributed to classification performance, with a focus on the presynaptic vesicle cloud and the postsynaptic density.

Finally, SynEM is sufficiently computationally efficient to be applied to large connectomics datasets. The total runtime on the 384592  $\mu\text{m}^3$  dataset was 2.6 hours on a mid-size computational cluster (480 CPU cores, 16GB RAM per core). This would imply a runtime of 279.9 days for a large 1  $\text{mm}^3$  dataset, which is comparable to the time required for current segmentation methods, but much faster than the currently required human annotation time ( $10^5$  to  $10^6$  h, Fig. 1c). Note that SynEM was not yet optimized for computational speed (plain matlab code, see Suppl. Code and git repository posted at <https://gitlab.mpcdf.mpg.de/connectomics/SynEM>).

275

## 276 **SynEM for connectomes**

277

278 We so far evaluated SynEM on the basis of the detection performance of single  
279 synaptic interfaces. Since we are interested in measuring the connectivity matrices of  
280 large-scale mammalian cortical circuits (connectomes) we obtained a statistical  
281 estimate of connectome error rates based on synapse detection error rates. We  
282 assume that the goal is a binary connectome containing the information whether  
283 pairs of neurons are connected or not. Automated synapse detection provides us  
284 with weighted connectomes reporting the number of synapses between neurons,  
285 from which we can obtain binary connectomes by considering all neuron pairs with at  
286 least  $\gamma_{nn}$  synapses as connected (Fig. 5a). Synaptic connections between neurons in  
287 the mammalian cerebral cortex have been found to be established via multiple  
288 synapses per neuron pair (Fig. 5b, Feldmeyer et al., 1999, Feldmeyer et al., 2002,  
289 Feldmeyer et al., 2006, Frick et al., 2008, Markram et al., 1997, range 1-8 synapses  
290 per connection, mean  $4.3 \pm 1.4$  for excitatory connections). The effect of synapse  
291 recall  $R_s$  on recall of neuron-to-neuron connectivity  $R_{nn}$  can be estimated (Fig. 5c) for  
292 each threshold  $\gamma_{nn}$  given the distribution of the number of synapses per connected  
293 neuron pair  $n_{syn}$ . For connectomes in which neuron pairs with at least one detected  
294 synapse are considered as connected ( $\gamma_{nn} = 1$ ), a neuron-to-neuron connectivity  
295 recall  $R_{nn}$  of 97% can be achieved with a synapse detection recall  $R_s$  of 65.1% (Fig.  
296 5c, black arrow) if synapse detection is independent between multiple synapses of  
297 the same neuron pair. SynEM achieves 99.4% synapse detection precision  $P_s$  at this  
298 recall (Fig. 3e).

299 The resulting precision of neuron-to-neuron connectivity  $P_{nn}$  then follows from the  
300 total number of synapses in the connectome  $N_{syn} = N^2 \times c_r \times \langle n_{syn} \rangle$ , with  $c_r$  the pairwise  
301 connectivity rate, about 20% for local excitatory connections in cortex (Feldmeyer et  
302 al., 1999),  $\langle n_{syn} \rangle$  the mean number of synapses per connection ( $4.3 \pm 1.4$ , Fig. 5b),  
303 and  $N^2$  the size of the connectome. A fraction  $R_s$  of these synapses is detected (true  
304 positive detections, TPs). The number of false positive (FP) synapse detections was  
305 deduced from TP and the synapse precision  $P_s$  as  $FP = TP \times (1 - P_s) / P_s$ , yielding  
306  $R_s \times N_{syn} \times (1 - P_s) / P_s$  false positive synapse detections. These we assumed to be  
307 distributed randomly on the connectome and estimated how often at least  $\gamma_{nn}$   
308 synapses fell into a previously empty connectome entry. These we considered as  
309 false positive connectome entries, whose rate yields the binary connectome

precision  $P_{nn}$  (see Methods for details of the calculation). At  $R_{nn}$  of 97.1%, SynEM yields a neuron-to-neuron connection precision  $P_{nn}$  of 98.5% (Fig. 5d, black arrow, Fig. 5e; note that this result is stable against varying underlying connectivity rates  $c_{re}=5\%..30\%$ , see indicated ranges in Fig. 5e).

For the treatment of inhibitory connections, we followed the notion that synapse detection performance could be optimized by restricting classifications to interfaces established by inhibitory axons (as we had analogously seen for restricting analysis to spine synapses above, Fig. 3e). For this, we evaluated SynEM on a test set of inhibitory axons for which we classified all neurite contacts of these axons (171 synapses, 9430 interfaces). While the precision and recall for single inhibitory synapses is lower than for excitatory ones (75% recall, 82% precision, Fig. 5 – figure supplement 1, SynEM<sup>(i)</sup><sub>s</sub>), the higher number of synapses per connected cell pair ( $n_{syn}^{(i)}$  is on average about 6, Suppl. File 3, Gupta et al., 2000; Markram et al., 2004; Koelbl et al., 2015; Hoffmann et al., 2015) still yields substantial neuron-to-neuron precision and recall also for inhibitory connectomes (98% recall, 97% precision, Fig. 5e, Fig. 5 – figure supplement 1, SynEM<sup>(i)</sup><sub>nn</sub>; this result is stable against varying underlying inhibitory connectivity rates  $c_{ri}=20\%..80\%$ , see ranges indicated in Fig. 5e). Error rates of less than 3% for missed connections and for wrongly detected connections are well below the noise of synaptic connectivity so far found in real biological circuits (e.g., Helmstaedter et al., 2013, Bartol et al., 2015), and thus likely sufficient for a large range of studies involving the mapping of cortical connectomes.

In summary, SynEM provides fully automated detection of synapses, their synaptic partner neurites and synapse direction for binary mammalian connectomes up to 97% precision and recall, a range which was previously prohibitively expensive to attain in large-scale volumes by existing methods (Fig. 5e, Fig. 5 – figure supplement 2).

## Local cortical connectome

We applied SynEM to a sparse local cortical connectome between 104 axons and 100 postsynaptic processes in the dataset from L4 of mouse cortex (Fig. 6a, neurites were reconstructed using webKnossos (Boergens et al., 2017) and SegEM as previously reported (Berning et al., 2015)). We first detected all contacts and calculated the total contact area between each pair of pre- and postsynaptic

processes (“contactome”, Fig. 6b). We then classified all contacts using SynEM (at the classification threshold  $\theta_{nn}$  (Table 3) yielding 98.5% precision and 97.1% recall for excitatory neuron-to-neuron connections and 97.3% precision and 98.5% recall for inhibitory neuron-to-neuron connections) to obtain the weighted connectome  $C_w$  (Fig. 6c). The detected synapses were clustered when they were closer than 1500 nm for a given neurite pair. This allowed us to concatenate large synapses with multiple active zones or multiple contributing SegEM segments into one (Fig. 6 – figure supplement 1). To obtain the binary connectome we thresholded the weighted connectome at  $\gamma_{nn} = 1$  for excitatory and at  $\gamma_{nn} = 2$  for inhibitory neuron-to-neuron connections (Fig. 6d). The resulting connectome contained 880 synapses distributed over 536 connections.

### Frequency and size of automatically detected synapses

Finally, to check whether SynEM-detected synapses matched previous reports on synapse frequency and size, we applied SynEM to half of the entire cortex dataset used for this study (i.e. a volume of 192296  $\mu\text{m}^3$ ). SynEM detected 195644 synapses, i.e. a synapse density of 1.02 synapses per  $\mu\text{m}^3$ , consistent with previous reports (Merchan-Perez et al., 2014).

We then measured the size of the axon-spine interface of SynEM detected synapses in the test set (Fig. 7a, b). We find axon-spine interface size of  $0.263 \pm 0.206 \mu\text{m}^2$  (mean  $\pm$  s.d.; range 0.033 – 1.189  $\mu\text{m}^2$ ;  $n = 181$ ), consistent with previous reports (de Vivo et al., 2017: (SW)  $0.297 \pm 0.297 \mu\text{m}^2$  ( $p = 0.518$ , two-sample two-tailed t-test on the natural logarithm of the axon-spine interface size), (EW)  $0.284 \pm 0.275 \mu\text{m}^2$  ( $p = 0.826$ , two-sample two-tailed t-test on the natural logarithm of the axon-spine interface size)). This indicates that, first, synapse detection in our lower-resolution SBEM data (in-plane image resolution about 11 nm, section thickness about 26-30 nm) yields similar synapse size distributions as in the higher-resolution data in de Vivo et al., 2017 (in-plane image resolution 5.9 nm; section thickness about 50 nm) and, secondly, that SynEM-based synapse detection has no obvious bias towards larger synapses.

376

## 377 **DISCUSSION**

378

379 We report SynEM, a toolset for automated synapse detection in EM-based  
380 connectomics. The particular achievement is that the synapse detection for densely  
381 mapped connectomes from the mammalian cerebral cortex is fully automated  
382 yielding below 3% residual error in the binary connectome. Importantly, SynEM  
383 directly provides the location and size of synapses, the involved neurites and the  
384 synapse direction without human interaction. With this, synapse detection is removed  
385 as a bottleneck in large-scale mammalian connectomics.

386

387 Evidently, synapse detection is facilitated in high-resolution EM data, and becomes  
388 most feasible in FIB-SEM data at a resolution of about 4-8 nm isotropic (Kreshuk et  
389 al., 2011, Fig. 3f). Yet, only by compromising resolution for speed (and thus volume)  
390 of imaging, the mapping of large, potentially even whole-brain connectomes is  
391 becoming plausible (Fig. 3f). Therefore it was essential to obtain automated synapse  
392 detection for EM data that is of lower resolution and scalable to such volumes. The  
393 fact that SynEM also outperforms state-of-the-art methods on high-resolution  
394 anisotropic 3D EM data (Fig. 3f, Roncal et al., 2015) indicates that our approach of  
395 segmentation-based interface classification has merits in a wider range of 3D EM  
396 data modalities.

397

398 In addition to high image resolution, recently proposed special fixation procedures  
399 that enhance the extracellular space in 3D EM data (Pallotto et al., 2015) are  
400 reported to simplify synapse detection for human annotators. In such data, direct  
401 touch between neurites has a very high predictive power for the existence of a  
402 (chemical or electrical) synapse, since otherwise neurite boundaries are separated  
403 by extracellular space. Thus, it is expected that such data also substantially simplifies  
404 automated synapse detection. The advantage of SynEM is that it achieves fully  
405 automated synapse detection in conventionally stained and fixated 3D EM data, in  
406 which neurite contact is most frequent at non-synaptic sites. Such data is widely  
407 used, and acquiring such data does not require special fixation protocols.

408

409 Finally, our approach to selectively classify interfaces of inhibitory axons (Fig. 5f, Fig  
410 5 – figure supplement 1) requires discussion. So far, the classification of synapses



411 into inhibitory (symmetric) vs. excitatory (asymmetric) was carried out for a given single  
412 synapse, often in single cross sections of single synapses (e.g. Colonnier, 1968).  
413 With the increasing availability of large-scale 3D EM datasets, however, synapse  
414 types can be defined based on multiple synapses of the same axon (e.g. Kasthuri et  
415 al., 2015). In the case of a dataset sized a cubic millimeter of cortical tissue, most  
416 axons of interneurons will be fully contained in the dataset since most inhibitory  
417 neurons are local. Consequently, the classification of single synapses can be  
418 replaced by the assignment of synapses to the respective axon; the type of axon is  
419 then inferred from the neurons' somatic and dendritic features. Even for axons which  
420 are not completely contained in the dataset, the assignment to inhibitory or excitatory  
421 synaptic phenotypes can be based on dozens or hundreds rather than single  
422 synapses.

423  
424 Together, SynEM resolves synapse detection for high-throughput cortical  
425 connectomics of mammalian brains, removing synapse detection as a bottleneck in  
426 connectomics. With this, SynEM renders the further acceleration of neurite  
427 reconstruction again the key challenge for future connectomic analysis.

## METHODS

### Annotation time estimates

Neuropil composition (Fig. 1b) was considered as follows: Neuron density of 157,500 per  $\text{mm}^3$  (White & Peters, 1993), axon path length density of 4 km per  $\text{mm}^3$  and dendrite path length density of 1 km per  $\text{mm}^3$  (Braitenberg & Schüz, 1998), spine density of about 1 per  $\mu\text{m}$  dendritic shaft length, with about 2  $\mu\text{m}$  spine neck length per spine (thus twice the dendritic path length), synapse density of 1 synapse per  $\mu\text{m}^3$  (Merchan-Perez et al., 2014) and bouton density of 0.1 – 0.25 per  $\mu\text{m}$  axonal path length (Braitenberg & Schüz, 1998). Annotation times were estimated as 200 - 400 h per mm path length for contouring, 3.7 – 7.2 h/mm path length for skeletonization (Helmstaedter et al., 2011, Helmstaedter et al., 2013, Berning et al., 2015), 0.6 h/mm for flight-mode annotation (Boergens et al., 2017), 0.1 h/ $\mu\text{m}^3$  for synapse annotation by volume search (estimated from the test set annotation) and an effective interaction time of 60 s per identified bouton for axon-based synapse search. All annotation times refer to single-annotator work hours, redundancy may be increased to reduce error rates in neurite and synapse annotation in these estimates (see Helmstaedter et al., 2011).

### EM image dataset and segmentation

SynEM was developed and tested on a SBEM dataset from layer 4 of mouse primary somatosensory cortex (dataset 2012-09-28\_ex145\_07x2, K.M.B. and M.H., unpublished data, see also Berning et al., 2015). Tissue was conventionally en-bloc stained (Briggman et al., 2011) with standard chemical fixation yielding compressed extracellular space (compare to Pallotto et al., 2015).

The image dataset was volume segmented using the SegEM algorithm (Berning et al., 2015). Briefly, SegEM was run using CNN 20130516T204040<sub>8,3</sub> and segmentation parameters as follows:  $r_{\text{se}} = 0$ ;  $\theta_{\text{ms}} = 50$ ;  $\theta_{\text{hm}} = 0.39$ ; (see last column in Table 2 in (Berning et al., 2015)). For training data generation, a different voxel threshold for watershed marker size  $\theta_{\text{ms}} = 10$  was used. For test set and local connectome calculation the SegEM parameter set optimized for whole cell segmentations was used ( $r_{\text{se}} = 0$ ;  $\theta_{\text{ms}} = 50$ ;  $\theta_{\text{hm}} = 0.25$ , see Table 2, Berning et al., 2015).

## Neurite interface extraction and subvolume definition

Interfaces between a given pair of segments in the SegEM volume segmentation were extracted by collecting all voxels from the one-voxel boundary of the segmentation for which that pair of segments was present in the boundary's 26-neighborhood. Then, all interface voxels for a given pair of segments were linked by connected components, and if multiple connected components were created, these were treated as separate interfaces. Interface components with a size of 150 voxels or less were discarded.

To define the subvolumes around an interface used for feature aggregation (Fig. 2b), we collected all voxels that were at a maximal distance of 40, 80 and 160 nm from any interface voxel and that were within either of the two adjacent segments of the interface. The interface itself was also considered as a subvolume yielding a total of 7 subvolumes for each interface.

## Feature calculation

Eleven 3-dimensional image filters with one to 15 instances each (Table 1) were calculated as follows and aggregated over the 7 subvolumes of an interface using 9 summary statistics, yielding 3224 features per directed interface. Image filters were applied to cuboids of size 548x548x268 voxels, each, which overlapped by 72,72 and 24 voxels in x,y and z dimension, respectively, to ensure that all interface subvolumes were fully contained in the filter output.

Gaussian filters were defined by evaluating the unnormalized 3d Gaussian density function

$$\hat{g}_{\sigma}(x, y, z) = \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2} - \frac{z^2}{2\sigma_z^2}\right)$$

at integer coordinates  $(x, y, z) \in U = \{-f_x, -f_x-1, \dots, f_x\} \times \{-f_y, -f_y-1, \dots, f_y\} \times \{-f_z, -f_z-1, \dots, f_z\}$  for a given standard deviation  $\sigma = (\sigma_x, \sigma_y, \sigma_z)$  and a filter size  $f = (f_x, f_y, f_z)$  and normalizing the resulting filter by the sum over all its elements

$$g_{\sigma}(x, y, z) = \frac{\hat{g}_{\sigma}(x, y, z)}{\sum_{(x', y', z') \in U} \hat{g}_{\sigma}(x', y', z')}.$$

First and second order derivatives of Gaussian filters were defined as

$$\frac{\partial}{\partial x} g_{\sigma}(x, y, z) = g_{\sigma}(x, y, z) \frac{-x}{\sigma_x^2},$$

$$\frac{\partial^2}{\partial x^2} g_{\sigma}(x, y, z) = g_{\sigma}(x, y, z) \left( \frac{x^2}{\sigma_x^2} - 1 \right) \frac{1}{\sigma_x^2},$$

$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} g_{\sigma}(x, y, z) = g_{\sigma}(x, y, z) \frac{xy}{\sigma_x^2 \sigma_y^2}.$$

494 and analogously for the other partial derivatives. Normalization of  $g_{\sigma}$  and evaluation  
 495 of derivatives of Gaussian filters was done on U as described above. Filters were  
 496 applied to the raw data I via convolution (denoted by  $*$ ) and we defined the image's  
 497 Gaussian derivatives as

$$I_x^{\sigma}(x, y, z) = I * \frac{\partial g_{\sigma}}{\partial x}(x, y, z),$$

$$I_{xy}^{\sigma}(x, y, z) = I * \frac{\partial^2 g_{\sigma}}{\partial x \partial y}(x, y, z)$$

498 and analogously for the other partial derivatives.

499 Gaussian smoothing was defined as  $I * g_{\sigma}$ .

500 Difference of Gaussians was defined as  $(I * g_{\sigma} - I * g_{k\sigma})$ , where the standard deviation of  
 501 the second Gaussian filter is multiplied element-wise by the scalar k.

502 Gaussian gradient magnitude was defined as

$$\sqrt{I_x^{\sigma}(x, y, z)^2 + I_y^{\sigma}(x, y, z)^2 + I_z^{\sigma}(x, y, z)^2}.$$

503 Laplacian of Gaussian was defined as

$$I_{xx}^{\sigma}(x, y, z) + I_{yy}^{\sigma}(x, y, z) + I_{zz}^{\sigma}(x, y, z)$$

504 Structure tensor S was defined as a matrix of products of first order Gaussian  
 505 derivatives, convolved with an additional Gaussian filter (window function)  $g_{\sigma_w}$ :

$$S_{xy} = (I_x^{\sigma_D} I_y^{\sigma_D}) * g_{\sigma_w}$$

506 and analogously for the other dimensions, with standard deviation  $\sigma_D$  of the image's  
 507 Gauss derivatives. Since S is symmetric, only the diagonal and upper diagonal

508 entries were determined, the eigenvalues were calculated and sorted by increasing  
509 absolute value.

510 The Hessian matrix was defined as the matrix of second order Gaussian derivatives:

$$H_{xy} = I_{xy}^{\sigma},$$

511 and analogously for the other dimensions. Eigenvalues were calculated as described  
512 for the Structure tensor.

513 The local entropy feature was defined as

$$- \sum_{L \in \{0, \dots, 255\}} p(L) \log_2 p(L),$$

514 where  $p(L)$  is the relative frequency of the voxel intensity in the range  $\{0, \dots, 255\}$  in a  
515 given neighborhood  $U$  of the voxel of interest (calculated using the `entropyfilt` function  
516 in MATLAB).

517 Local standard deviation for a voxel at location  $(x, y, z)$  was defined by

$$\sqrt{\frac{1}{|U| - 1} \sum_{(x', y', z') \in U} I(x', y', z') - \frac{1}{|U|(|U| - 1)} \left( \sum_{(x', y', z') \in U} I(x', y', z') \right)^2},$$

518 for the neighborhood  $U$  of location  $(x, y, z)$  with  $|U|$  number of elements and  
519 calculated using MATLAB's `stdfilt` function.

520 Sphere average was defined as the mean raw data intensity for a spherical  
521 neighborhood  $U_r$  with radius  $r$  around the voxel of interest, with

$$U_r = \{(x, y, z) | x^2 + y^2 + (2z)^2 \leq r^2\} \cap Z^3,$$

522 where  $Z^3$  is the 3 dimensional integer grid;  $x, y, z$  are voxel indices;  $z$  anisotropy was  
523 approximately corrected.

524 The intensity/variance feature for voxel location  $(x, y, z)$  was defined as

$$\sum_{(x', y', z') \in U} I(x', y', z')^2 - \left( \sum_{(x', y', z') \in U} I(x', y', z') \right)^2,$$

525 for the neighborhood  $U$  of location  $(x, y, z)$ .

The set of parameters for which filters were calculated is summarized in Table 1.

11 shape features were calculated for the border subvolume and the two 160 nm-restricted subvolumes, respectively. For this, the center locations (midpoints) of all voxels of a subvolume were considered. Shape features were defined as follows: The number of voxel feature was defined as the total number of voxels in the subvolumes. The voxel based diameter was defined as the diameter of a sphere with the same volume as the number of voxels of the subvolumes. Principal axes lengths were defined as the three eigenvalues of the covariance matrix of the respective voxel locations. Principal axes product was defined as the scalar product of the first principal components of the voxel locations in the two 160 nm-restricted subvolumes. Voxel based convex hull was defined as the number of voxels within the convex hull of the respective subvolume voxels (calculated using the `convhull` function in MATLAB).

## **Generation of training and validation labels**

Interfaces were annotated by 3 trained undergraduate students using a custom-written GUI (in MATLAB, Fig. 3 – figure supplement 1). A total of 40 non-overlapping rectangular volumes within the center  $86 \times 52 \times 86 \mu\text{m}^3$  of the dataset were selected (39 sized  $5.6 \times 5.6 \times 5.6 \mu\text{m}^3$  each and one of size  $9.6 \times 6.8 \times 8.3 \mu\text{m}^3$ ). Then, all interfaces within these volumes were extracted as described above. Interfaces with a center of mass less than  $1.124 \mu\text{m}$  from the volume border were not considered. For each interface, a raw data volume of size  $(1.6 \times 1.6 \times 0.7\text{--}1.7) \mu\text{m}^3$ , centered on the center of mass of the interface voxel locations was presented to the annotator. When the center of mass was not part of the interface, the closest interface voxel was used. The raw data was rotated such that the second and third principal components of the interface voxel locations (restricted to a local surround of  $15 \times 15 \times 7$  voxels around the center of mass of the interface) defined the horizontal and vertical axes of the displayed images. First, the image plane located at the center of mass of the interface was shown. The two segmentation objects were transparently overlaid (Fig. 3 – figure supplement 1) in separate colors (the annotator could switch the labels off for better visibility of raw data). The annotator had the option to play a video of the image stack or to manually browse through the images. The default video playback started at the first image. An additional video playback mode started at the center of mass of the interface, briefly transparently highlighted the segmentation objects of

the interface, and then played the image stack in reverse order to the first plane and from there to the last plane. In most cases, this already yielded a decision. In addition, annotators had the option to switch between the 3 orthogonal reslices of the raw data at the interface location (Fig. 3 – figure supplement 1). The annotators were asked to label the presented interfaces as non-synaptic or synaptic. For the synaptic label, they were asked to indicate the direction of the synapse (see Fig. 3 – figure supplement 1). In addition to the annotation label interfaces could be marked as “undecided”. Interfaces were annotated by one annotator each. The interfaces marked as undecided were validated by an expert neuroscientist. In addition, all synapse annotations were validated by an expert neuroscientist, and a subset of non-synaptic interfaces was cross-checked. Together, 75,383 interfaces (1858 synaptic, 73,525 non-synaptic) were labeled this way. Of these, the interfaces from 8 label volumes (391 synaptic and 11906 non-synaptic interfaces) were used as validation set; the interfaces from the other 32 label volumes were used for training.

## **SynEM classifier training and validation**

The target labels for the undirected, augmented and directed label sets were defined as described in the Results (Fig. 3c). We used boosted decision stumps (level-one decision trees) trained by the AdaBoostM1 (Freund & Schapire, 1997) or LogitBoost (Friedman et al., 2000) implementation from the MATLAB Statistical Toolbox (fitensemble). In both cases the learning rate was set to 0.1 and the total number of weak learners to 1500. Misclassification cost for the synaptic class was set to 100. Precision and recall values of classification results were reported with respect to the synaptic class. For validation, the undirected label set was used, irrespective of the label set used in training. If the classifier was trained using the directed label set then the thresholded prediction for both orientations were combined by logical OR.

## **Test set generation and evaluation**

To obtain an independent test set disjunct from the data used for training and validation, we randomly selected a volume of size 512 x 512 x 256 voxels ( $5.75 \times 5.75 \times 7.17 \mu\text{m}^3$ ) from the dataset that contained no soma or dominantly large dendrite. One volume was not used because of unusually severe local image

alignment issues which are meanwhile solved for the entire dataset. The test volume had the bounding box [3713, 2817, 129, 4224, 3328, 384] in the dataset. First, the volume was searched for synapses (see Fig. 1d) in webKnossos (Boergens et al., 2017) by an expert neuroscientist. Then, all axons in the volume were skeleton-traced using webKnossos. Along the axons, synapses were searched (strategy in Fig. 1e) by inspecting vesicle clouds for further potential synapses. Afterwards the expert searched for vesicle clouds not associated with any previously traced axon and applied the same procedure as above. In total, that expert found 335 potential synapses. A second expert neuroscientist used the tracings and synapse annotations from the first expert to search for further synapse locations. The second expert added 8 potential synapse locations. All 343 resulting potential synapses were collected and independently assessed by both experts as synaptic or not. The experts labeled 282 potential locations as synaptic, each. Of these, 261 were in agreement. The 42 disagreement locations (21 from each annotator) were re-examined jointly by both experts and validated by a third expert on a subset of all synapses. 18 of the 42 locations were confirmed as synaptic, of which one was just outside the bounding box. Thus, in total, 278 synapses were identified. The precision and recall of the two experts in their independent assessment with respect to this final set of synapses was 93.6%, 94.6% (expert 1) and 97.9%, 98.9% (expert 2), respectively.

Afterwards all shaft synapses were labeled by the first expert and proofread by the second. Subsequently, the synaptic interfaces were voxel-labeled to be compatible with the method by Becker et al. This initial test set comprised 278 synapses, of which 36 were labeled as shaft/inhibitory.

Next, all interfaces between pairs of segmentation objects in the test volume were extracted as described above. Then, the synapse labels were assigned to those interfaces whose border voxels had any overlap with one of the 278 voxel-labeled synaptic interfaces. Afterwards, these interface labels were again proof-read by an expert neuroscientist. Finally, interfaces closer than 160 nm from the boundary of the test volume were excluded to ensure that interfaces were fully contained in the test volume. The final test set comprised 235 synapses out of which 31 were labeled as shaft/inhibitory. With this we obtained a high-quality test set providing both voxel-labeled synapses and synapse labels for interfaces, to allow the comparison of different detection methods.



For the calculation of precision and recall, a synapse was considered detected if at least one interface that had overlap with the synapse was detected by the classifier (TPs); a synapse was considered missed if no overlapping interface of a given synapse was detected (FNs); and a detection was considered false positive (FP) if the corresponding interface did not overlap with any labeled synapse.

## **Inhibitory synapse detection**

The labels for inhibitory-focused synapse detection were generated using skeleton tracings of inhibitory axons. Two expert neuroscientists used these skeleton tracings to independently detect all synapse locations along the axons. Agreeing locations were considered synapses and disagreeing locations were resolved jointly by both annotators. The resulting test set contains 171 synapses. Afterwards, all SegEM segments of the consensus postsynaptic neurite were collected locally at the synapse location. For synapse classification all interfaces in the dataset were considered that contained one SegEM segment located in one of these inhibitory axons. Out of these interfaces all interfaces were labeled synaptic that were between the axon and a segment identified as postsynaptic. The calculation of precision and recall curves was done as for the dense test set (see above) by considering a synapse detected if at least one interface overlapping with it was detected by the classifier (TPs); a synapse was considered missed if no interface of a synapse was detected (FNs); and a detection was considered false positive (FP) if the corresponding interface did not overlap with any labeled synapse.

## **Comparison to previous work**

The approach of Becker et al., 2012 was evaluated using the implementation provided in Ilastik (Sommer et al., 2011). This approach requires voxel labels of synapses. We therefore first created training labels: an expert neuroscientist created sparse voxel labels at interfaces between pre- and postsynaptic processes and twice as many labels for non-synaptic voxels for five cubes of size  $3.4 \times 3.4 \times 3.4 \mu\text{m}^3$  that were centered in five of the volumes used for training SynEM. Synaptic labels were made for 115 synapses (note that the training set in Becker et al., 2012 only contained 7-20 synapses). Non-synaptic labels were made for two training cubes

first. The non-synaptic labels of the remaining cubes were made in an iterative fashion by first training the classifier on the already created synaptic and non-synaptic voxel labels and then adding annotations specifically for misclassified locations using Ilastik. Eventually, non-synaptic labels in the first two training cubes were extended using the same procedure.

For voxel classification all features proposed in (Becker et al., 2012) and 200 weak learners were used. The classification was done on a tiling of the test set into cubes of size  $256 \times 256 \times 256$  voxels ( $2.9 \times 2.9 \times 7.2 \mu\text{m}^3$ ) with a border of 280 nm around each tile. After classification, the borders were discarded, and tiles were stitched together. The classifier output was thresholded and morphologically closed with a cubic structuring element of three voxels edge length. Then, connected components of the thresholded classifier output with a size of at least 50 voxels were identified. Synapse detection precision and recall rates were determined as follows: A ground truth synapse (from the final test set) was considered detected (TP) if it had at least a single voxel overlap with a predicted component. A ground truth synapse was counted as a false negative detection if it did not overlap with any predicted component (FN). To determine false positive classifications, we evaluated the center of the test volume (shrunk by 160 nm from each side to  $484 \times 484 \times 246$  voxels) and counted each predicted component that did not overlap with any of the ground truth synapses as false positive detection (FP). For this last step, we used all ground truth synapses from the initial test set, in favor of the Becker et al. classifier.

For comparison with (Kreshuk et al., 2014) the same voxel training data as for (Becker et al., 2012) was used. The features provided by Ilastik up to a standard deviation of 5 voxels for the voxel classification step were used. For segmentation of the voxel probability output map the graph cut segmentation algorithm of Ilastik was used with label smoothing ( $[1, 1, 0.5]$  voxel standard deviation), a voxel probability threshold of 0.5 and graph cut constant of  $\lambda = 0.25$ . Objects were annotated in five additional cubes of size  $3.4 \times 3.4 \times 3.4 \mu\text{m}^3$  that were centered in five of the interface training set cubes different from the one used for voxel prediction resulting in 299 labels (101 synaptic, 198 non-synaptic). All object features provided by Ilastik were used for object classification. The evaluation on the test set was done as for (Becker et al., 2012).

For comparison with (Dorkenwald et al., 2017) six of the 32 training cubes used for interface classification with a total volume of  $225 \mu\text{m}^3$  were annotated with voxel labels for synaptic junctions, vesicle clouds and mitochondria. The annotation of

vesicle clouds and mitochondria was done using voxel predictions of a convolutional neural network (CNN) trained on mitochondria, vesicle clouds and membranes. The membrane predictions were discarded and the vesicle clouds and mitochondria labels were first proofread by undergraduate students and then twice by an expert neuroscientist. The voxels labels for synaptic junctions were added by an expert neuroscientist based on the identified synapses in the interface training data. Overall 310 synapses were annotated in the training volume. A recursive multi-class CNN was trained on this data with the same architecture and hyperparameter settings as described in (Dorkenwald et al., 2017) using the ElektroNN framework. For the evaluation of synapse detection performance only the synaptic junction output was used. The evaluation on the test set was done as for (Becker et al., 2012) with a connected component threshold of 250 voxels.

## **Evaluation on the dataset from Kasthuri et al., 2015**

The image data, neurite and synapse segmentation from (Kasthuri et al., 2015) hosted on openconnecto.me (kasthuri11cc, kat11segments, kat11synapses) was used (downloaded using the provided scripts at <https://github.com/neurodata-arxiv/CAJAL> ). The segmentation in the bounding box [2432, 7552; 6656, 10112; 769, 1537] (resolution 1) was adapted to have a one-voxel boundary between segments by first morphologically eroding the original segmentation with a 3-voxel cubic structuring element and running the MATLAB watershed function on the distance-transform of the eroded segmentation on a tiling with cubes of size [1024, 1024, 512] voxels. Since the Kasthuri et al., 2015 segmentation in the selected bounding box was not dense, voxels with a segment id of zero in the original segmentation whose neighbors at a maximal distance of 2 voxels (maximum-distance) also all had segment ids zero were set to segment id zero in the adapted segmentation. All segments in the adapted segmentation that were overlapping with a segment in the original segmentation were set to the id of the segment in the original segmentation. The bounding box [2817, 6912; 7041, 10112; 897, 1408] of the resulting segmentation was tiled into non-overlapping cubes of [512, 512, 256] voxels. For all synapses in the synapse segmentation the pre- and postsynaptic segment of the synapse were marked using webKnossos (Boergens et al., 2017) and all interfaces between the corresponding segments at a maximal distance of 750 nm to the synapse centroid that were also overlapping with an object in the synapse

segmentation were associated to the corresponding synapse and assigned a unique group id. Only synapses labeled as “sure” in Kasthuri et al., 2015 were evaluated. All interfaces with a center of mass in the region ac3 with the bounding box [5472, 6496; 8712, 9736; 1000, 1256] were used for testing. All interfaces with a center of mass at a distance of at least 1  $\mu\text{m}$  to ac3 were used for training if there was no interface between the same segment ids in the test set. Interfaces between the same segment ids as an interface in the test set were only considered for training if the distance to ac3 was above 2  $\mu\text{m}$ . For feature calculation the standard deviation of Gaussian filters was adapted to the voxel size 6 x 6 x 30 nm of the data (i.e.  $\sigma$  in Table 2 was set to 12/2 in x- and y-dimension and 12/30 in z-dimension). The directed label set approach was used for classification. The calculation of precision recall rates was done as described above (“test set generation and evaluation”).

## Pairwise connectivity model

The neuron-to-neuron connection recall was calculated assuming an empirical distribution  $p(n)$  of the number of synapses  $n$  between connected excitatory neurons given by published studies (see Supp. Table 2, Feldmeyer et al., 1999, Feldmeyer et al., 2002, Feldmeyer et al., 2006, Frick et al., 2008, Markram et al., 1997). For inhibitory connections we used a fixed value of 6 synapses (see Supp. Table 3, Koelbl et al., 2015, Hoffmann et al., 2015, Gupta et al., 2000, Markram et al., 2004). We further assumed that the number of retrieved synapses is given by a binomial model with retrieval probability given by the synapse classifier recall  $R_s$  on the test set:

$$P(k \geq \gamma_{nn} | R_s) = \sum_n \text{Bin}(k \geq \gamma_{nn} | n, R_s) p(n),$$

Where  $\gamma_{nn}$  is the threshold on the number of synapses between a neuron pair to consider it as connected (see Fig. 5a). This equates to the neuron-to-neuron recall:  $R_{nn} = P(k \geq \gamma_{nn} | R_s)$ .

To compute the neuron-to-neuron precision, we first calculated the expected number of false positive synapse detections ( $\text{FP}_s$ ) made by a classifier with precision  $P_s$  and recall  $R_s$ :

$$FP_s = \frac{(1 - P_s)}{P_s} R_s N_{syn}$$

763 where  $N_{syn}$  is the total number of synapses in a dataset calculated from the average  
 764 number of synapses per connected neuron pair  $\langle n_{syn} \rangle$  times the number of  
 765 connected neuron pairs  $N_{con}$  and  $c_r$  is the connectivity ratio given by  $N_{con}/N^2$  with  $N$   
 766 the number of neurons in the connectome.

767 We then assumed that these false positive synapse detections occur randomly and  
 768 therefore are assigned to one out of  $N^2$  possible neuron-to-neuron connections with a  
 769 frequency  $FP_s/N^2$ .

770 We then used a Poisson distribution to estimate the number of cases in which at  
 771 least  $\gamma_{nn}$   $FP_s$  synapses would occur in a previously zero entry of the connectome,  
 772 yielding a false positive neuron-to-neuron connection ( $FP_{nn}$ ).

$$FP_{nn} = N^2 (1 - c_r) Poi(x \geq \gamma_{nn} | FP_s/N^2).$$

773 Finally, the true positive detections of neuron-to-neuron connections in the  
 774 connectome  $TP_{nn}$  are given in terms of the neuron-to-neuron connection recall  $R_{nn}$  by

$$TP_{nn} = N^2 c_r R_{nn}.$$

775 Together, the neuron-to-neuron connection precision  $P_{nn}$  is given by

$$P_{nn} = \frac{TP_{nn}}{TP_{nn} + FP_{nn}} = \frac{c_r R_{nn}}{c_r R_{nn} + (1 - c_r) Poi(x \geq \gamma_{nn} | FP_s/N^2)}.$$

776 The connectivity ratio was set to  $c_r = 0.2$  (Feldmeyer et al., 1999) for excitatory and to  
 777 0.6 for inhibitory connections (Gibson et al., 1999, Koelbl et al., 2015).

778

## 779 **Local connectome**

780

781 For determining the local connectome (Fig. 6) between 104 pre- and 100  
 782 postsynaptic processes, we used 104 axonal skeleton tracings (traced at 1 to 5-fold  
 783 redundancy) and 100 dendrite skeleton tracings. 10 axons were identified as  
 784 inhibitory and are partially contained in the inhibitory test set. All volume objects  
 785 which overlapped with any of the skeleton nodes were detected and concatenated to  
 786 a given neurite volume. Then, all interfaces between pre- and postsynaptic  
 787 processes were classified by SynEM. The area of each interface was calculated as in

(Berning et al., 2015) and the total area of all contacts between all neurite pairs was calculated (Fig. 6b). To obtain the weighted connectome  $C_w$  (Fig. 6c), we applied the SynEM scores threshold  $\theta_{nn}$  (Table 3) for the respective presynaptic type (excitatory, inhibitory). Detected synaptic interfaces were clustered using hierarchical clustering (single linkage, distance cutoff 1,500 nm) if the interfaces were between the same pre- and postsynaptic objects. To obtain the binary connectome  $C_{bin}$  (Fig. 6d) we thresholded the weighted connectome at the connectome threshold  $\gamma_{nn} = 1$  for excitatory and  $\gamma_{nn} = 2$  for inhibitory connections (Table 3). The overall number of synapses in the dataset was calculated by considering all interfaces above the score threshold for the best single synapse performance ( $\theta_s$ ) as synaptic. To obtain the final synapse count the retrieved synaptic interfaces were clustered using hierarchical clustering with single linkage and a distance cutoff between the centroids of the interfaces of 320.12 nm (this distance cutoff was obtained by optimizing the synapse density prediction on the test set).

### **Axon-spine interface area comparison**

For the evaluation of axon-spine interface area (ASI) all spine synapses in the test set were considered for which SynEM had detected at least one overlapping neurite interface (using  $\theta_s$  for spine synapses, Fig. 3e). The ASI of a detected synapse was calculated by summing the area of all interfaces between segmentation objects that overlapped with the synapse. For comparison to ASI distributions obtained at higher imaging resolution in a recent study (spontaneous wake (SW) and enforced wake (EW) conditions reported in Table S1 in de Vivo et al., 2017), it was assumed that the ASI distributions are lognormal (see de Vivo et al., 2017, Fig. 2B). Two-sample two-tailed t-tests were performed for comparing the natural logarithmic values of the SynEM-detected ASI from the test set ( $\log \text{ASI} -1.60 \pm 0.74$ ,  $n=181$ ; mean  $\pm$  s.d.) with the lognormal distributions for SW and EW from de Vivo et al., 2017, ( $\log \text{ASI} -1.56 \pm 0.83$ ,  $n=839$ , SW;  $-1.59 \pm 0.81$ ,  $n=836$ , EW; mean  $\pm$  s.d.),  $p = 0.5175$  (SW) and  $p = 0.8258$  (EW).

819

820

821 **Code and data availability**

822

823 All code used to train and run SynEM is available under the MIT license in the  
824 Supplementary Code and will be made available at  
825 <https://gitlab.mpcdf.mpg.de/connectomics/SynEM> upon publication. To run SynEM,  
826 please follow instructions in the readme.md file in Suppl. Code. Data used to train  
827 and evaluate SynEM will be made available at <https://synem.rzg.mpg.de/webdav/>.

828

829

## ACKNOWLEDGEMENTS

We thank Jan Gleixner for first test experiments on synapse detection and fruitful discussions in an early phase of the project, Alessandro Motta for comments on the manuscript, Christian Guggenberger for excellent support with compute infrastructure, Raphael Jacobi, Raphael Kneissl, Athanasia Natalia Marahori, and Anna Satzger for data annotation and Elias Eulig, Robin Hesse, Martin Schmidt, Christian Schramm and Matej Zecevic for data curation. We thank Heiko Wissler and Dalila Rustemovic for support with illustrations.



## FIGURE LEGENDS

### Figure 1

The challenge of synapse detection in connectomics. **(a)** Sketch of mouse primary somatosensory cortex (S1) with circuit modules (“barrels”) in cortical layer 4 and minimum required dataset extent for a “barrel” dataset (250  $\mu\text{m}$  edge length) and a dataset extending over the whole cortical depth from pia to white matter (WM) (1 mm edge length). **(b)** Number of synapses and neurons, total axonal, dendritic and spine path length for the example datasets in (a) (White & Peters, 1993, Braitenberg & Schüz, 1998, Merchan-Perez et al., 2014). **(c)** Reconstruction time estimates for neurites and synapses; For synapse search strategies see sketches in d,e. Dashed arrows: latest skeletonization tools (webKnossos, Boergens et al., 2017) allow for a further speed up of neurite skeletonization by about 5-to-10-fold, leaving synapse detection as the main annotation bottleneck. **(d)** Volume search for synapses by visually investigating 3d image stacks and keeping track of already inspected locations takes about 0.1 h/ $\mu\text{m}^3$ . **(e)** Axon-based synapse detection by following axonal processes and detecting synapses at boutons consumes about 1 min per bouton. **(f)** Examples of synapses imaged at an in-plane voxel size of 6 nm and **(g)** 12 nm in conventionally en-bloc stained and fixated tissue (Briggman et al., 2011, Hua et al., 2015) imaged using SBEM (Denk & Horstmann, 2004). Arrows: synapse locations. Note that synapse detection in high-resolution data is much facilitated in the plane of imaging. Large-volume image acquisition is operated at lower resolution, requiring better synapse detection algorithms. **(h)** Synapse shown in 3D EM raw data, resliced in the 3 orthogonal planes. Scale bars in f and h, 500 nm. Scale bar in f applies to g.

## Figure 2

Synapse detection by classification of neurite interfaces. **(a)** Definition of interfaces used for synapse classification in SynEM. Raw EM data (left) is first volume segmented (using SegEM, Berning et al., 2015). Neighboring volume segments are identified (right). **(b)** Definition of perisynaptic subvolumes used for synapse classification in SynEM consisting of a border (red) and subvolumes adjacent to the neurite interface extending to distances of 40, 80 and 160 nm. **(c)** Example outputs of two texture filters: the difference of Gaussians (DoG) and the intensity/variance filter (int./var.). Note the clear signature of postsynaptic spine heads (right). **(d)** Distributions of int/var. texture filter output for image voxels at a synaptic (top) and non-synaptic interface (bottom). Medians over subvolumes are indicated (arrows, color scale as in b). **(e)** SynEM flow chart. Scale bars, 500 nm. Scale bar in a applies to a,b.

### Figure 3

SynEM training and evaluation. **(a)** Histogram of SynEM scores calculated on the validation set. Fully automated synapse detection is obtained by thresholding the SynEM score at threshold  $\theta$ . **(b)** SynEM scores for the two possible directions of interfaces. Note that SynEM scores are disjunct in a threshold regime used for best single synapse performance ( $\theta_s$ ) and best neuron-to-neuron recall and precision ( $\theta_{nn}$ ), see Fig. 5, indicating a clear bias towards one of the two possible synaptic directions. **(c)** Strategy for label generation. Based on annotator labels (Ann. Label), three types of label sets were generated: Initial label set ignored interface orientation (Undir.); Augmented label set included mirror-reflected interfaces (Augment.); Directed label set used augmented data but considered only one synaptic direction as synaptic (Directed, see also Fig. 3 – figure supplement 1). **(d)** Development of the SynEM classifier. Classification performance for different features, aggregation statistics, classifier parameters and label sets. Init: initial classifier used (see Table 1). The initial classifier was extended by using additional features (Add feat, see Table 1, first row), 40 and 80 nm subvolumes for feature aggregation (Add subvol, see Fig. 2b) and aggregate statistics (Add stats, see Table 1). Direct: Classifier trained on directed label set (see Fig. 3c). Logit: Classifier trained on full feature space using LogitBoost. Augment & Logit: Logit classifier trained on augmented label set (see Fig. 3c). Direct & Logit: Logit classifier trained on directed label set (see Fig. 3c). **(e)** Test set performance on 3D SBEM data of SynEM (purple) evaluated for spine and shaft synapses (all synapses, solid line) and for spine synapses (exc. synapses, dashed line), only. Threshold values for optimal single synapse detection performance (black circle) and an optimal connectome reconstruction performance (black square, see Fig. 5). (see also Fig. 3 – figure supplement 2) **(f)** Relation between 3D EM imaging resolution, imaging speed and 3D EM experiment duration (top), exemplified for a dataset sized 1 mm<sup>3</sup>. Note that the feasibility of experiments strongly depends on the chosen voxel size. Bottom: published synapse detection performance (reported as F1 score) in dependence of the respective imaging resolution (see also Suppl. File 1). dark blue, Mishchenko et al., 2010; cyan, Kreshuk et al., 2011; light gray, Becker et al., 2012; dark gray, Kreshuk et al., 2014; red, Roncal et al., 2015; green, Dorkenwald et al., 2017; Black brackets indicate direct comparison of SynEM to top-performing methods: SynEM vs Roncal et al., 2015 on ATUM-SEM dataset (Kasthuri et al., 2015); SynEM vs Dorkenwald et al., 2017 and Becker et al., 2012 on our test set. See Fig. 3 – figure supplement 3 for comparison of Precision-Recall curves. Note that SynEM outperforms the previously top-

performing methods. Note also that most methods provide synapse detection, but require the detection of synaptic partners and synapse direction in a separate classification step. Gray solid line: drop of partner detection performance compared to synapse detection in Dorkenwald et al., 2017; dashed gray lines, analogous possible range of performance drop as reported for bird dataset in Dorkenwald et al., 2017. SynEM combines synapse detection and partner detection into one classification step.

## Figure 4

SynEM classification and feature importance. **(a)** SynEM classification examples at  $\theta_s$  (circle in e). True positive (TP), true negative (TN), False negative (FN) and false positive (FP) interface classifications (blue arrow, classified interface) shown as 3 image planes spaced by 56 nm (i.e. every second SBEM data slice, top to bottom). Note that synapse detection in 3D SBEM data requires inspection of typically 10-20 consecutive image slices (see Synapse Gallery in Supplementary File 4 for examples). 1: presynaptic; 2: postsynaptic; x: non-synaptic. Note for the FP example that the axonal bouton (1) innervates a neighboring spine head, but the interface to the neurite under classification (x) is non-synaptic (blue arrow). **(b)** Ranked classification importance of SynEM features. All features (top left), relevance of feature quality (bottom left), subvolumes (top right) and pooling statistics (bottom right). Note that only 378 features contribute to classification. See Table 3 for the 10 feature instances of highest importance, Table 1 for feature name abbreviations, and text for details. Scale bars, 500 nm.

942

943 **Figure 5**

944 Effect of SynEM classification performance on error rates in automatically mapped  
945 binary connectomes. **(a)** Sketch of a weighted connectome (left) reporting the  
946 number of synapses per neuron-to-neuron connection, transformed into a binary  
947 connectome (middle) by considering neuron pairs with at least  $\gamma_{nn}$  synapses as  
948 connected. **(b)** Distribution of reported synapse number for connected excitatory  
949 neuron pairs obtained from paired recordings in rodent cerebral cortex (Feldmeyer et  
950 al., 1999, Feldmeyer et al., 2002, Feldmeyer et al., 2006, Frick et al., 2008, Markram  
951 et al., 1997). Average distribution (cyan) is used for the precision estimates in the  
952 following (see Suppl. File 2). **(c)** Relationship between SynEM recall for single  
953 interfaces (synapses)  $R_s$  and the ensuing neuron-to-neuron connectome recall  $R_{nn}$   
954 (recall in  $C_{bin}$ , a) for each of the excitatory cortico-cortical connections (summarized  
955 in b) and for connectome binarization thresholds of  $\gamma_{nn} = 1$  and  $\gamma_{nn} = 2$  (full and  
956 dashed, respectively). **(d)** Relationship between SynEM precision for single  
957 interfaces (synapses)  $P_s$  and the ensuing neuron-to-neuron connectome precision  
958  $P_{nn}$ . Colors as in c. (for inhibitory synapses see also Fig. 5 – figure supplement 1) **(e)**  
959 Predicted remaining error in the binary connectome (reported as 1-F1 score for  
960 neuron-to-neuron connections) for fully automated synapse classification using  
961 SynEM on 3D EM data from mouse cortex using two different imaging modalities:  
962 ATUM-SEM (left, Kasthuri et al., 2015) and our data using SBEM (right). e,i:  
963 excitatory or inhibitory connectivity model (see b and methods) shown for  $c_{re}=20\%$   
964 and  $c_{ri}=60\%$ . Black lines indicate range for varying assumptions of pairwise  
965 connectivity rate  $c_{re} = (5\%, 10\%, 30\%)$  (excitatory) and  $c_{ri} = (20\%, 40\%, 80\%)$   
966 (inhibitory). Note that SynEM yields a remaining error of close to or less than 2%,  
967 well below expected biological wiring noise, allowing for fully automated synapse  
968 detection in large-scale binary connectomes. See Suppl. Fig. 5 – figure supplement 2  
969 for comparison to previous synapse detection methods.

970

## Figure 6

Example sparse local cortical connectome obtained using SynEM. **(a)** 104 axonal (94 excitatory, 10 inhibitory) and 100 dendritic processes within a volume sized  $86 \times 52 \times 86 \mu\text{m}^3$  from layer 4 of mouse cortex skeletonized using webKnossos (Boergens et al., 2017), volume segmented using SegEM (Berning et al., 2015). **(b)** Contactome reporting total contact area between pre- and postsynaptic processes. **(c)** Weighted connectome obtained at the SynEM threshold  $\theta_{nn}$  optimized for the respective presynaptic type (excitatory, inhibitory) (see Fig 2e, black square, Table 3). (see also Fig. 6 – figure supplement 1) **(d)** Binary connectome obtained from the weighted connectome by thresholding at  $\gamma_{nn} = 1$  for excitatory connections and  $\gamma_{nn} = 2$  for inhibitory connections. The resulting predicted neuron-to-neuron recall and precision were 98%, 98% for excitatory and 98%, 97% for inhibitory connections, respectively (see Fig. 5e). Green: number of pre- (right) and postsynaptic (bottom) partners for each neurite.

## Figure 7

Comparison of synapse size in SBEM data. (a) Distribution of axon-spine interface area ASI for the SynEM-detected synapses onto spines in the test set from mouse S1 cortex imaged at  $11.24 \times 11.24 \times 28 \text{ nm}^3$  voxel size (see Fig. 3e), purple; and distributions from de Vivo et al., 2017 in S1 cortex from mice under two wakefulness conditions (SW: spontaneous wake, EW: enforced wake), imaged at higher resolution of  $5.9 \text{ nm}$  (xy plane) with a section thickness of  $54.7 \pm 4.8 \text{ nm}$  (SW),  $51.4 \pm 10.3 \text{ nm}$  (EW) (de Vivo et al., 2017). (b) Same distributions as in (a) shown on natural logarithmic scale ( $\log \text{ ASI SynEM } -1.60 \pm 0.74, n=181$ ;  $\log \text{ ASI SW } -1.56 \pm 0.83, n=839$ ;  $\log \text{ ASI EW } -1.59 \pm 0.81, n=836$ ; mean  $\pm$  s.d.). Note that the distributions are indistinguishable ( $p=0.52$  (SynEM vs. SW),  $p=0.83$  (SynEM vs. EW), two-sample two-tailed t-test), indicating that the size distribution of synapses detected in our lower-resolution data is representative, and that SynEM does not have a substantial detection bias towards larger synapses.



## REFERENCES

- Bartol, T. M., Bromer, C., Kinney, J., Chirillo, M. A., Bourne, J. N., Harris, K. M., & Sejnowski, T. J. (2015). Nanoconnectomic upper bound on the variability of synaptic plasticity. *Elife*, 4, e10778. doi:10.7554/eLife.10778
- Becker, C., Ali, K., Knott, G., & Fua, P. (2012). *Learning context cues for synapse segmentation in EM volumes*. Paper presented at the International Conference on Medical Image Computing and Computer-Assisted Intervention.
- Becker, C., Ali, K., Knott, G., & Fua, P. (2013). Learning context cues for synapse segmentation. *IEEE Trans Med Imaging*, 32(10), 1864-1877. doi:10.1109/TMI.2013.2267747
- Berning, M., Boergens, K. M., & Helmstaedter, M. (2015). SegEM: Efficient Image Analysis for High-Resolution Connectomics. *Neuron*, 87(6), 1193-1206. doi:10.1016/j.neuron.2015.09.003
- Boergens, K. M., Berning, M., Bocklisch, T., Braunlein, D., Drawitsch, F., Frohnhofer, J., . . . Helmstaedter, M. (2017). webKnossos: efficient online 3D data annotation for connectomics. *Nat Methods*. doi:10.1038/nmeth.4331
- Braitenberg, V., & Schüz, A. (1998). *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media.
- Briggman, K. L., Helmstaedter, M., & Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471(7337), 183-188. doi:10.1038/nature09818
- Colonnier, M. (1968). Synaptic patterns on different cell types in the different laminae of the cat visual cortex. An electron microscope study. *Brain Res*, 9(2), 268-287.
- de Vivo, L., Bellesi, M., Marshall, W., Bushong, E. A., Ellisman, M. H., Tononi, G., & Cirelli, C. (2017). Ultrastructural evidence for synaptic scaling across the wake/sleep cycle. *Science*, 355(6324), 507-510. doi:10.1126/science.aah5982
- Denk, W., & Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biol*, 2(11), e329. doi:10.1371/journal.pbio.0020329
- Dorkenwald, S., Schubert, P. J., Killinger, M. F., Urban, G., Mikula, S., Svara, F., & Kornfeld, J. (2017). Automated synaptic connectivity inference for volume electron microscopy. *Nat Methods*, 14(4), 435-442. doi:10.1038/nmeth.4206
- Eberle, A. L., Mikula, S., Schalek, R., Lichtman, J., Knothe Tate, M. L., & Zeidler, D. (2015). High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *J Microsc*, 259(2), 114-120. doi:10.1111/jmi.12224
- Feldmeyer, D., Egger, V., Lubke, J., & Sakmann, B. (1999). Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single 'barrel' of developing rat somatosensory cortex. *J Physiol*, 521 Pt 1, 169-190.
- Feldmeyer, D., Lubke, J., & Sakmann, B. (2006). Efficacy and connectivity of intracolumnar pairs of layer 2/3 pyramidal cells in the barrel cortex of juvenile rats. *J Physiol*, 575(Pt 2), 583-602. doi:10.1113/jphysiol.2006.105106
- Feldmeyer, D., Lubke, J., Silver, R. A., & Sakmann, B. (2002). Synaptic connections between layer 4 spiny neurone-layer 2/3 pyramidal cell pairs in juvenile rat barrel cortex: physiology and anatomy of interlaminar signalling within a cortical column. *J Physiol*, 538(Pt 3), 803-822.

- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119-139. doi:DOI 10.1006/jcss.1997.1504
- Frick, A., Feldmeyer, D., Helmstaedter, M., & Sakmann, B. (2008). Monosynaptic connections between pairs of L5A pyramidal neurons in columns of juvenile rat somatosensory cortex. *Cereb Cortex*, 18(2), 397-406. doi:10.1093/cercor/bhm074
- Friedman, J., Hastie, T., & Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *Annals of Statistics*, 28(2), 337-374. doi:DOI 10.1214/aos/1016218223
- Gibson, J. R., Beierlein, M., & Connors, B. W. (1999). Two networks of electrically coupled inhibitory neurons in neocortex. *Nature*, 402(6757), 75-79. doi:10.1038/47035
- Gray, E. G. (1959). Axo-somatic and axo-dendritic synapses of the cerebral cortex: an electron microscope study. *J Anat*, 93, 420-433.
- Gupta, A., Wang, Y., & Markram, H. (2000). Organizing principles for a diversity of GABAergic interneurons and synapses in the neocortex. *Science*, 287(5451), 273-278.
- Hayworth, K., Kasthuri, N., Schalek, R., & Lichtman, J. (2006). Automating the collection of ultrathin serial sections for large volume TEM reconstructions. *Microscopy and Microanalysis*, 12(S02), 86-87.
- Helmstaedter, M. (2013). Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nat Methods*, 10(6), 501-507. doi:10.1038/nmeth.2476
- Helmstaedter, M., Briggman, K. L., & Denk, W. (2011). High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nat Neurosci*, 14(8), 1081-1088. doi:10.1038/nn.2868
- Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., & Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500(7461), 168-174. doi:10.1038/nature12346
- Hoffmann, J. H., Meyer, H. S., Schmitt, A. C., Straehle, J., Weitbrecht, T., Sakmann, B., & Helmstaedter, M. (2015). Synaptic Conductance Estimates of the Connection Between Local Inhibitor Interneurons and Pyramidal Neurons in Layer 2/3 of a Cortical Column. *Cereb Cortex*, 25(11), 4415-4429. doi:10.1093/cercor/bhv039
- Hua, Y., Laserstein, P., & Helmstaedter, M. (2015). Large-volume en-bloc staining for electron microscopy-based connectomics. *Nat Commun*, 6, 7923. doi:10.1038/ncomms8923
- Huang, G. B., Scheffer, L. K., & Plaza, S. M. (2016). Fully-Automatic Synapse Prediction and Validation on a Large Data Set. *arXiv preprint arXiv:1604.03075*.
- Jagadeesh, V., Anderson, J., Jones, B., Marc, R., Fisher, S., & Manjunath, B. S. (2014). Synapse classification and localization in Electron Micrographs. *Pattern Recognition Letters*, 43, 17-24. doi:10.1016/j.patrec.2013.06.001
- Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., . . . Lichtman, J. W. (2015). Saturated Reconstruction of a Volume of Neocortex. *Cell*, 162(3), 648-661. doi:10.1016/j.cell.2015.06.054
- Knott, G., Marchman, H., Wall, D., & Lich, B. (2008). Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *J Neurosci*, 28(12), 2959-2964. doi:10.1523/JNEUROSCI.3189-07.2008
- Koelbl, C., Helmstaedter, M., Lubke, J., & Feldmeyer, D. (2015). A barrel-related interneuron in layer 4 of rat somatosensory cortex with a high intrabarrel connectivity. *Cereb Cortex*, 25(3), 713-725. doi:10.1093/cercor/bht263
- Kreshuk, A., Funke, J., Cardona, A., & Hamprecht, F. A. (2015). *Who is talking to whom: synaptic partner detection in anisotropic volumes of insect brain*.

1107 Paper presented at the International Conference on Medical Image  
 1108 Computing and Computer-Assisted Intervention.  
 1109 Kreshuk, A., Koethe, U., Pax, E., Bock, D. D., & Hamprecht, F. A. (2014). Automated  
 1110 detection of synapses in serial section transmission electron microscopy  
 1111 image stacks. *PLoS One*, 9(2), e87351. doi:10.1371/journal.pone.0087351  
 1112 Kreshuk, A., Straehle, C. N., Sommer, C., Koethe, U., Cantoni, M., Knott, G., &  
 1113 Hamprecht, F. A. (2011). Automated detection and segmentation of synaptic  
 1114 contacts in nearly isotropic serial electron microscopy images. *PLoS One*,  
 1115 6(10), e24899. doi:10.1371/journal.pone.0024899  
 1116 Markram, H., Lubke, J., Frotscher, M., Roth, A., & Sakmann, B. (1997). Physiology  
 1117 and anatomy of synaptic connections between thick tufted pyramidal  
 1118 neurones in the developing rat neocortex. *J Physiol*, 500 ( Pt 2), 409-440.  
 1119 Markram, H., Toledo-Rodriguez, M., Wang, Y., Gupta, A., Silberberg, G., & Wu, C.  
 1120 (2004). Interneurons of the neocortical inhibitory system. *Nat Rev Neurosci*,  
 1121 5(10), 793-807. doi:10.1038/nrn1519  
 1122 Marquez Neila, P., Baumela, L., Gonzalez-Soriano, J., Rodriguez, J. R., DeFelipe, J.,  
 1123 & Merchan-Perez, A. (2016). A Fast Method for the Segmentation of Synaptic  
 1124 Junctions and Mitochondria in Serial Electron Microscopic Images of the  
 1125 Brain. *Neuroinformatics*. doi:10.1007/s12021-015-9288-z  
 1126 Merchan-Perez, A., Rodriguez, J. R., Gonzalez, S., Robles, V., Defelipe, J.,  
 1127 Larranaga, P., & Bielza, C. (2014). Three-dimensional spatial distribution of  
 1128 synapses in the neocortex: a dual-beam electron microscopy study. *Cereb*  
 1129 *Cortex*, 24(6), 1579-1588. doi:10.1093/cercor/bht018  
 1130 Mikula, S., Binding, J., & Denk, W. (2012). Staining and embedding the whole mouse  
 1131 brain for electron microscopy. *Nat Methods*, 9(12), 1198-1201.  
 1132 doi:10.1038/nmeth.2213  
 1133 Mikula, S., & Denk, W. (2015). High-resolution whole-brain staining for electron  
 1134 microscopic circuit reconstruction. *Nat Methods*, 12(6), 541-546.  
 1135 doi:10.1038/nmeth.3361  
 1136 Mishchenko, Y., Hu, T., Spacek, J., Mendenhall, J., Harris, K. M., & Chklovskii, D. B.  
 1137 (2010). Ultrastructural analysis of hippocampal neuropil from the  
 1138 connectomics perspective. *Neuron*, 67(6), 1009-1020.  
 1139 doi:10.1016/j.neuron.2010.08.014  
 1140 Pallotto, M., Watkins, P. V., Fubara, B., Singer, J. H., & Briggman, K. L. (2015).  
 1141 Extracellular space preservation aids the connectomic analysis of neural  
 1142 circuits. *Elife*, 4. doi:10.7554/eLife.08206  
 1143 Perez, A. J., Seyedhosseini, M., Deerinck, T. J., Bushong, E. A., Panda, S.,  
 1144 Tasdizen, T., & Ellisman, M. H. (2014). A workflow for the automatic  
 1145 segmentation of organelles in electron microscopy image stacks. *Front*  
 1146 *Neuroanat*, 8, 126. doi:10.3389/fnana.2014.00126  
 1147 Plaza, S. M., Parag, T., Huang, G. B., Olbris, D. J., Saunders, M. A., & Rivlin, P. K.  
 1148 (2014). Annotating synapses in large EM datasets. *arXiv preprint*  
 1149 *arXiv:1409.1801*.  
 1150 Roncal, W. G., Pekala, M., Kaynig-Fittkau, V., Kleissas, D. M., Vogelstein, J. T.,  
 1151 Pfister, H., . . . Hager, G. D. (2015). VESICLE: Volumetric Evaluation of  
 1152 Synaptic Interfaces using Computer Vision at Large Scale *Proceedings of the*  
 1153 *British Machine Vision Conference (BMVC)* (pp. 81.81-81.13): BMVA Press.  
 1154 Sommer, C., Straehle, C., Kothe, U., & Hamprecht, F. A. (2011). Ilastik: Interactive  
 1155 Learning and Segmentation Toolkit. *2011 8th IEEE International Symposium*  
 1156 *on Biomedical Imaging: From Nano to Macro*, 230-233.  
 1157 Staffler, B., Berning, M., Boergens, K. M., Gour, A., van der Smagt, P., &  
 1158 Helmstaedter, M. (2017). SynEM: Automated synapse detection for  
 1159 connectomics. *bioRxiv*.  
 1160 url:http://www.biorxiv.org/content/early/2017/01/22/099994

1161 White, E. L., & Peters, A. (1993). Cortical modules in the posteromedial barrel  
1162 subfield (Sml) of the mouse. *J Comp Neurol*, 334(1), 86-96.  
1163 doi:10.1002/cne.903340107

1164

1165

| Features                     | Kreshuk<br>et al.,<br>2011 | Becker<br>et al.,<br>2012 | Init.<br>Class. | Syn<br>EM | Parameters   | N of<br>instances<br>* |
|------------------------------|----------------------------|---------------------------|-----------------|-----------|--|------------------------|
| <b>Texture:</b>              |                            |                           |                 |           |  |                        |
| Raw data                     |                            | ✓                         | ✓               | ✓         | -  | 1                      |
| 3 EVs of Structure<br>Tensor | ✓                          | ✓                         | ✓               | ✓         | $(\sigma_w, \sigma_d) = \{(s,s), (s,2s), (2s,s), (2s,2s), (3s,3s)\}$ | 15                     |
| 3 EVs of Hessian             | ✓                          | ✓                         | ✓               | ✓         | $\sigma = \{s, 2s, 3s, 4s\}$   | 12                     |
| Gaussian Smoothing           | ✓                          |                           | ✓               | ✓         | $\sigma = \{s, 2s, 3s\}$   | 3                      |
| Difference of Gaussians      | ✓                          |                           |                 | ✓         | $(\sigma,k) = \{(s, 1.5), (s, 2), (2s, 1.5), (2s, 2), (3s, 1.5)\}$   | 5                      |
| Laplacian of Gaussian        | ✓                          | ✓                         | ✓               | ✓         | $\sigma = \{s, 2s, 3s, 4s\}$   | 4                      |
| Gauss Gradient Magn.         | ✓                          | ✓                         | ✓               | ✓         | $\sigma = \{s, 2s, 3s, 4s, 5s\}$                                     | 5                      |
| Local standard deviation     |                            |                           |                 | ✓         | $U = 1_{5 \times 5 \times 5}$  | 1                      |
| Int./var.                    |                            |                           |                 | ✓         | $U = \{1_{3 \times 3 \times 3}, 1_{5 \times 5 \times 5}\}$           | 2                      |
| Local entropy                |                            |                           |                 | ✓         | $U = 1_{5 \times 5 \times 5}$  | 1                      |
| Sphere average               |                            |                           |                 | ✓         | $r = \{3, 6\}$   | 2                      |
| <b>Shape:</b>                |                            |                           |                 |           |  |                        |
| Number of voxels             |                            |                           | ✓               | ✓         | Bo, 160  | 3                      |
| Diameter (vx based)          |                            |                           |                 | ✓         | Bo   | 1                      |
| Lengths of principal axes    |                            |                           |                 | ✓         | Bo   | 3                      |
| Principal axis product       |                            |                           |                 | ✓         | 160  | 1                      |
| Convex hull (vx based)       |                            |                           |                 | ✓         | Bo, 160  | 3                      |

1166

1167

**Table 1**

1168 **Overview of the classifier features used in SynEM, and comparison with**  
1169 **existing methods.** 11 3-dimensional texture filters employed at various filter  
1170 parameters given in units of standard deviation (s) of Gaussian filters (s was  
1171 12/11.24 voxels in x and y-dimension and 12/28 voxels in z-dimension, sizes of filters

1172 were set to  $\sigma/s*\text{ceil}(2*s)$ ). When structuring elements were used,  $1_{a \times b \times c}$  refers to a  
1173 matrix of size  $a \times b \times c$  filled with ones and  $r$  specifies the semi-principal axes of an  
1174 ellipsoid of length  $(r, r, r/2)$  voxels in  $x, y$  and  $z$ -dimension. All texture features are  
1175 pooled by 9 summary statistics (quantiles (0.25, 0.5, 0.75, 0, 1), mean, variance,  
1176 skewness, kurtosis, respectively) over the 7 subvolumes around the neurite interface  
1177 (see Fig. 2b). Shape features were calculated for three of the subvolumes: border  
1178 (Bo) and the 160 nm distant pre- and postsynaptic volumes (160). Init. Class: initial  
1179 SynEM classifier (see Fig. 3d for performance evaluation). N of instances: number of  
1180 feature instances per subvolume ( $n=7$ ) and aggregate statistic ( $n=9$ ). \*: Total number  
1181 of employed features is 63 times reported instances for texture features. For shape  
1182 features, the reported number is the total number of instances used, together yielding  
1183 3224 features total.

1184

| Rank | Feature                          | Parameters                          | Subvolume  | Aggregate statistic   |
|------|----------------------------------|-------------------------------------|------------|-----------------------|
| 1    | EVs of Struct. Tensor (largest)  | $\sigma_w = 2s$ ,<br>$\sigma_D = s$ | 160 nm, S1 | Median                |
| 2    | EVs of Struct. Tensor (smallest) | $\sigma_w = 2s$ ,<br>$\sigma_D = s$ | 160 nm, S1 | Median                |
| 3    | Local entropy                    | $U = 1_{5 \times 5 \times 5}$       | 160 nm, S2 | Variance              |
| 4    | Difference of Gaussians          | $\sigma = 3s$ ,<br>$k = 1.5$        | Border     | 25 <sup>th</sup> perc |
| 5    | Difference of Gaussians          | $\sigma = 2s$ ,<br>$k = 1.5$        | Border     | Median                |
| 6    | EVs of Struct. Tensor (middle)   | $\sigma_w = 2s$ ,<br>$\sigma_D = s$ | 40 nm, S2  | Min                   |
| 7    | Int./var.                        | $U = 1_{3 \times 3 \times 3}$       | Border     | 75 <sup>th</sup> perc |
| 8    | EVs of Struct. Tensor (largest)  | $\sigma_w = 2s$ ,<br>$\sigma_D = s$ | 80 nm, S1  | 25 <sup>th</sup> perc |
| 9    | Gauss gradient magnitude         | $\sigma = s$                        | 40 nm, S2  | 25 <sup>th</sup> perc |
| 10   | Principal axes length (2nd)      | -                                   | Border     | -                     |

**Table 2**

**SynEM features ranked by ensemble predictor importance.** See Fig. 4b and Methods for details. Note that two of the newly introduced features and one of the shape features had high classification relevance (Local entropy, Int./var., Principal axes length; cf. Table 1).

| Threshold score                | Single synapse $P_s/R_s$ | Neuron-to-neuron $P_{nn}/R_{nn}$ |                   |
|--------------------------------|--------------------------|----------------------------------|-------------------|
|                                |                          | $\gamma_{nn} = 1$                | $\gamma_{nn} = 2$ |
| $\theta_s = -1.67$<br>(exc)    | 88.5% / 88.1%            | 72.5% / 99.7%                    | 98.1% / 95.6%     |
| $\theta_{nn} = -0.08$<br>(exc) | 99.4% / 65.1%            | 98.5% / 97.1%                    | 100% / 83.4%      |
| $\theta_s = -2.06$<br>(inh)    | 82.1% / 74.9%            | 77.1% / 100%                     | 92.7% / 99.5%     |
| $\theta_{nn} = -1.58$<br>(inh) | 88.6% / 67.8%            | 84.7% / 99.9%                    | 97.3% / 98.5%     |

1192

1193

**Table 3**

1194 **SynEM score thresholds and associated precision and recall.** SynEM score  
1195 thresholds  $\theta$  chosen for optimized single synapse detection ( $\theta_s$ ) and optimized  
1196 neuron-to-neuron connection detection ( $\theta_{nn}$ ) with respective single synapse precision  
1197 ( $P_s$ ) and recall ( $R_s$ ) and estimated neuron-to-neuron precision and recall rates ( $P_{nn}$ ,  
1198  $R_{nn}$ , respectively) for connectome binarization thresholds of  $\gamma_{nn} = 1$  and  $\gamma_{nn} = 2$  (see  
1199 Fig. 5).

1200

1201



1202 **FIGURE SUPPLEMENTS**

1203

1204 Figure 3 – figure supplement 1

1205

1206 Figure 3 – figure supplement 2

1207

1208 Figure 3 – figure supplement 3

1209

1210 Figure 3 – figure supplement 4

1211

1212 Figure 5 – figure supplement 1

1213

1214 Figure 5 – figure supplement 2

1215

1216 Figure 6 – figure supplement 1

1217

## **FIGURE SUPPLEMENTS: Legends**

### **Figure 3 – figure supplement 1**

Graphical user interface (implemented in MATLAB) for efficient annotation of neurite interfaces as used for generating the training and validation labels. 3D image data is centered to the neurite interface and rotated such that the second and third principal components of the neurite interface span the displayed image plane. Segments are indicated by transparent overlay (interface, red; subsegment S1, blue and S2, green). Note that the test labels were independently annotated by volume search by multiple experts in webKnossos (Boergens et al., 2017), see Methods.

### **Figure 3 – figure supplement 2**

Distribution of training, validation and test data volumes within the dataset ex145\_07x2. Soma locations are indicated by spheres of radius 5  $\mu\text{m}$ .

### **Figure 3 – figure supplement 3**

Synapse detection performance comparison of SynEM with SyConn (Dorkenwald et al., 2017) and (Becker et al., 2012) on the 3D SBEM SynEM test set (Figure 3e). Note that while SynEM performs synapse detection and partner detection in one step these are separate steps in SyConn with an overall performance that is potentially different from the synapse detection step (in Dorkenwald et al., 2017, a reduction in performance by 10% in recall and 3% in precision from synapse detection to partner detection is reported, yielding a drop in F1 score of 0.057). Becker et al., 2012, does not contain a dedicated partner detection step.

### **Figure 3 – figure supplement 4**

Synapse detection performance comparison of SynEM with VesicleCNN (Dorkenwald et al., 2017; Roncal et al., 2015) on a 3D EM dataset from mouse S1 cortex obtained using ATUM-SEM (Kasthuri et al., 2015). Note that VesicleCNN was developed on that ATUM-SEM dataset.

#### **Figure 5 – figure supplement 1**

Performance of SynEM on a test set containing all interfaces between 3 inhibitory axons and all touching neurites (total of 9430 interfaces, 171 synapses). Single synapse detection precision and recall (solid line) and the ensuing predicted neuron-to-neuron precision and recall for inhibitory connections (dashed line) assuming on average 6 synapses for connections from interneurons (see Methods).

#### **Figure 5 – figure supplement 2**

Effect of synapse detection errors on predicted connectome error rates for competing methods. Predicted neuron-to-neuron errors (reported as  $(1 - F1 \text{ score})$  in percent) for the ATUM-SEM dataset (Kasthuri et al., 2015) using VesicleCNN (Roncal et al., 2015, orange) and for our SBEM dataset using Becker et al., 2012 (gray) and Syconn (Dorkenwald et al., 2017, green). Note that these approaches provide synapse detection, only. When including the detection of the synaptic partners, Dorkenwald et al., 2017 reported a drop of detection performance by 3% precision and 10% recall (indicated by gray crosses, tentatively also for the other approaches). SynEM provides synapse detection and partner detection together (compare to Fig. 5e).

#### **Figure 6 – figure supplement 1**

Procedure for obtaining synapse counts in the local connectome (Fig. 6). (a) Segmentation used for SynEM (note that a segmentation biased to neurite splits was used, see Berning et al., 2015) and (b) interfaces detected as synaptic (black lines). (c) combined skeleton-SegEM segmentation of neurites. (d) Synaptic neurite interfaces established between the same pre- and postsynaptic processes (as determined by the skeleton-SegEM segmentation, c) were clustered using hierarchical clustering with a distance cutoff of  $d = 1.5 \mu\text{m}$  (b) for obtaining the final synapse count. Scale bar, 500 nm.

## **SUPPLEMENTARY FILES**

### **Supplementary File (Table) 1**

**Overview of methods for automated synapse detection.** Res. Fac: Image voxel volume of SBEM data used in this study relative to the voxel volume in the reported studies. Note that most studies employ data of substantially higher image resolution.

### **Supplementary File (Table) 2**

**Number of synapses between connected neurons obtained from published studies of paired recordings of excitatory neurons in rodent cortex.** These distributions were used in Fig. 5 for prediction of connectome precision and recall.

### **Supplementary File (Table) 3**

**Number of synapses between connected neurons obtained from published studies of paired recordings of inhibitory neurons in rodent cortex.**

### **Supplementary File 4**

**Synapse gallery.** Document describing the criteria by which synapses in 3D SBEM data were detected by human expert annotators. These criteria are exemplified for synapses from the test set of the SynEM classifier.

## SOURCE DATA FILES

### **Figure 1 – source data 1**

Source data for plots in panels 1b, 1c

### **Figure 2 – source data 1**

Source data for plot in panel 2d

### **Figure 3 – source data 1**

Source data for plots in panels 3a, 3b, 3d, 3e, 3f

### **Figure 3 – figure supplement 3 – source data 1**

### **Figure 3 – figure supplement 4 – source data 1**

### **Figure 4 – source data 1**

Source data for plot in panel 4b

### **Figure 5 – source data 1**

Source data for plots in panels 5b, 5c, 5d, 5e

### **Figure 5 – figure supplement 1 – source data 1**

### **Figure 5 – figure supplement 2 – source data 1**

### **Figure 6 – source data 1**

Source data for plots in panels 6b, 6c, 6d

### **Figure 7 – source data 1**

Source data for plots in panels 7a, 7b

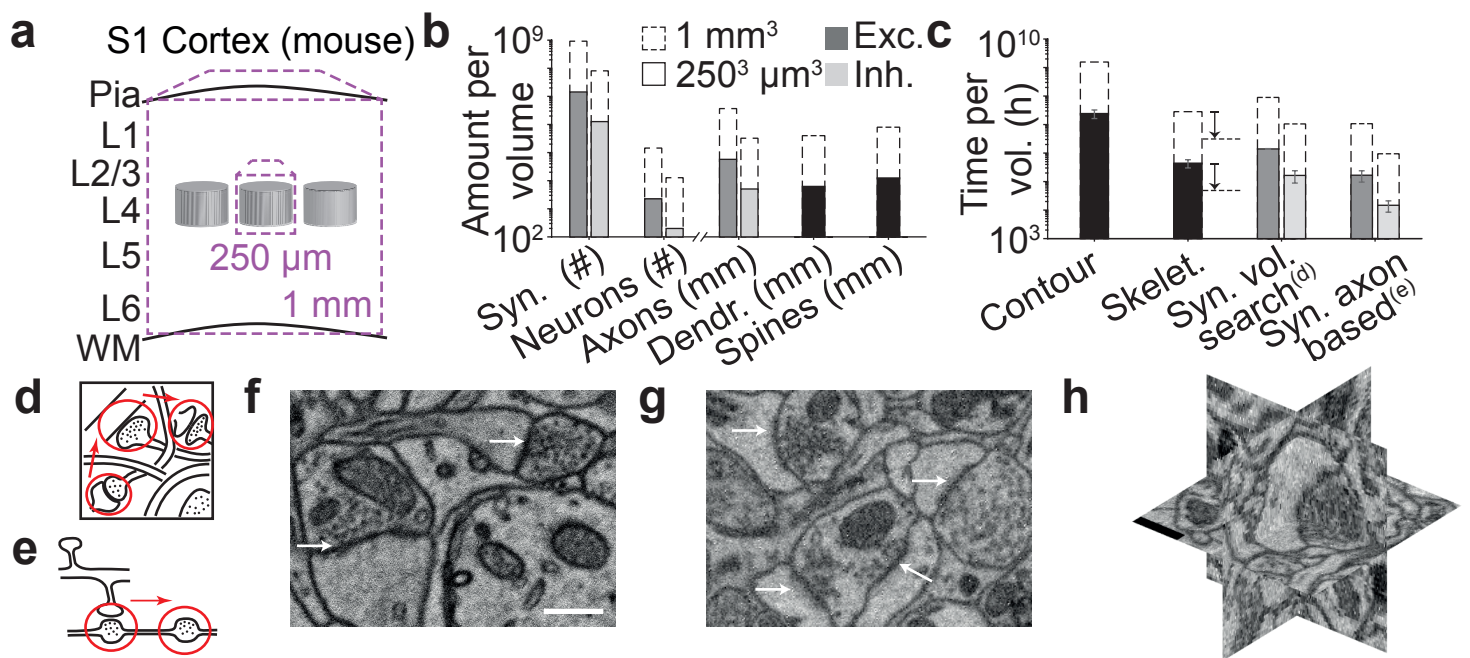


Figure 1  
Staffler et al.

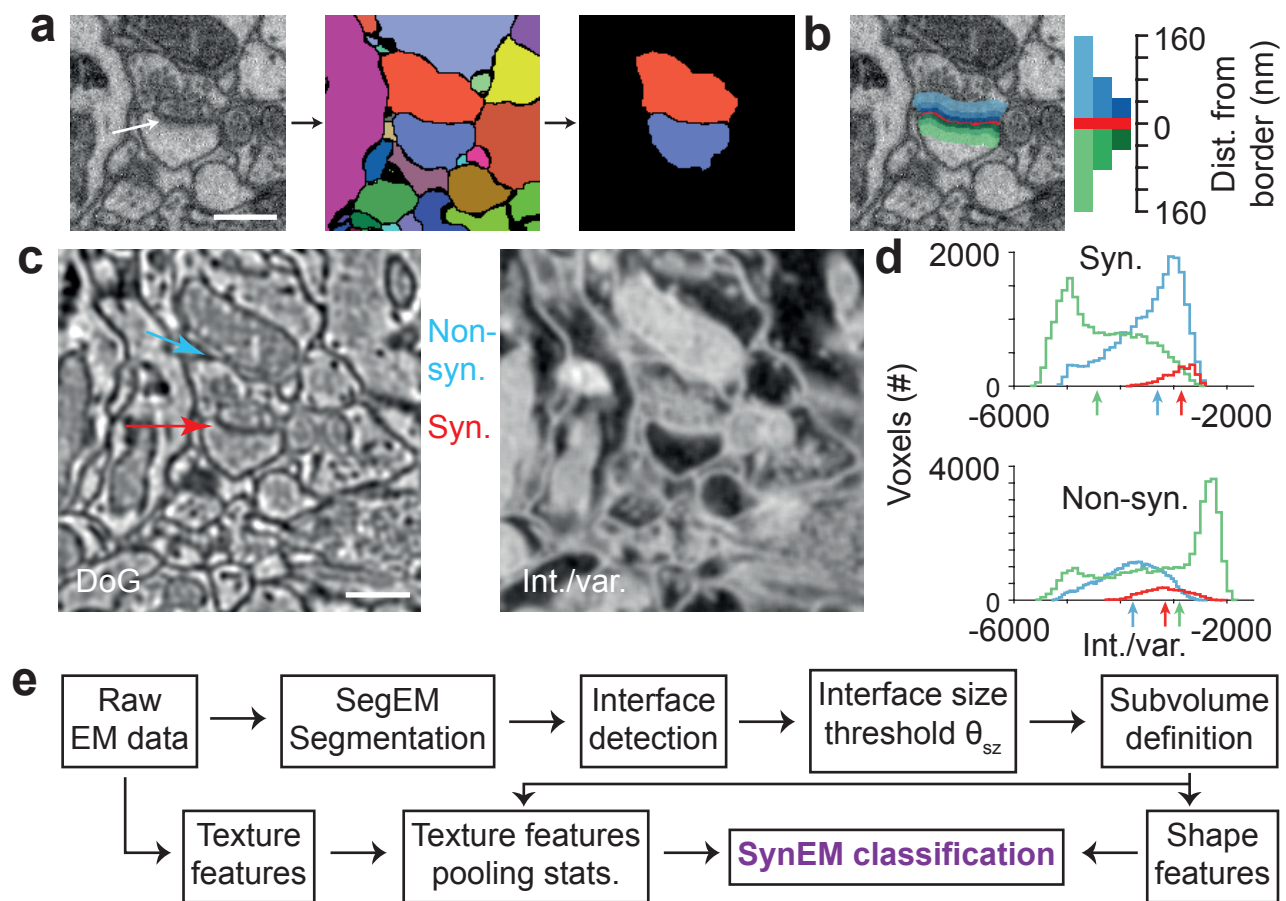


Figure 2  
Staffler et al.

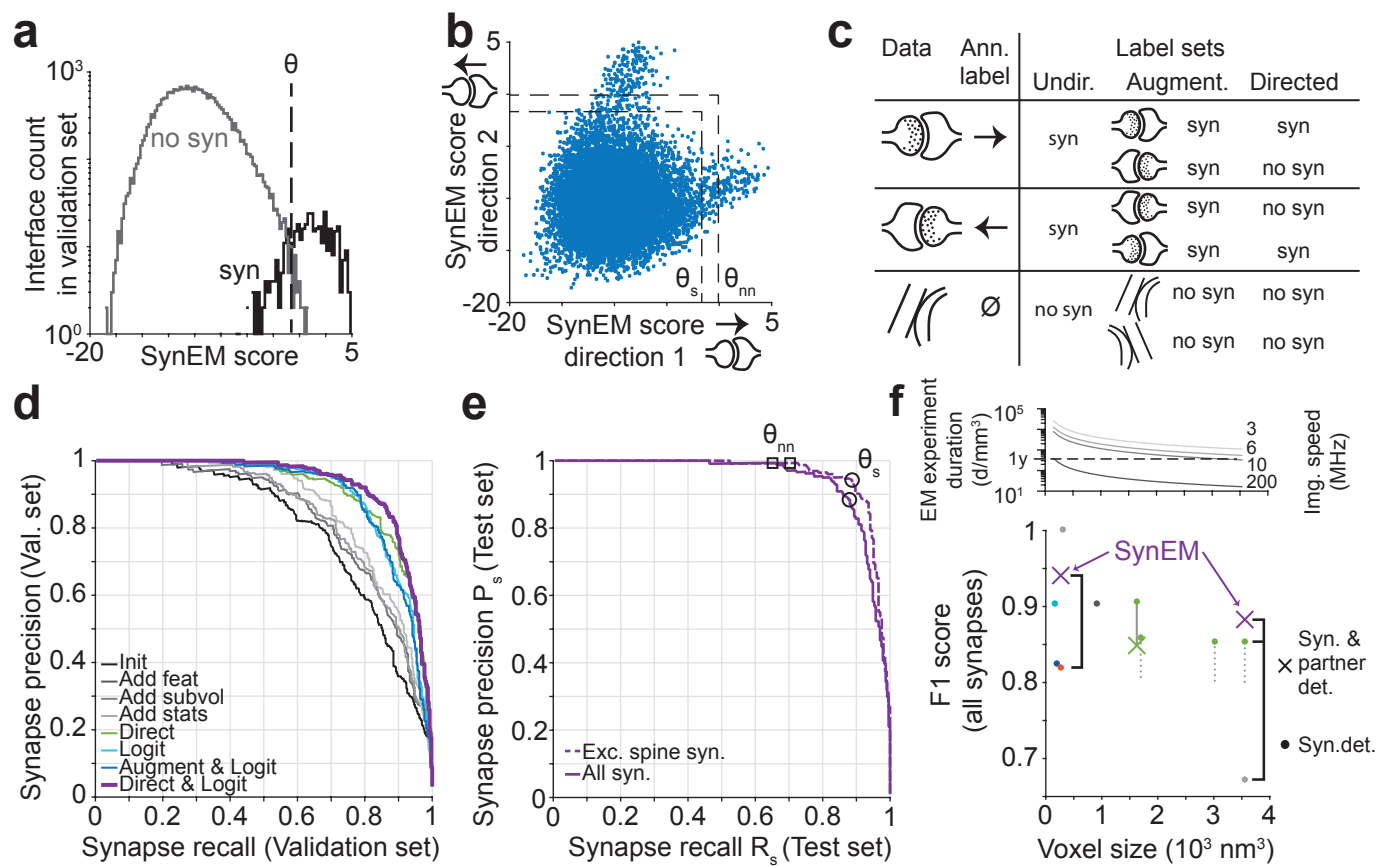


Figure 3  
Staffler et al.



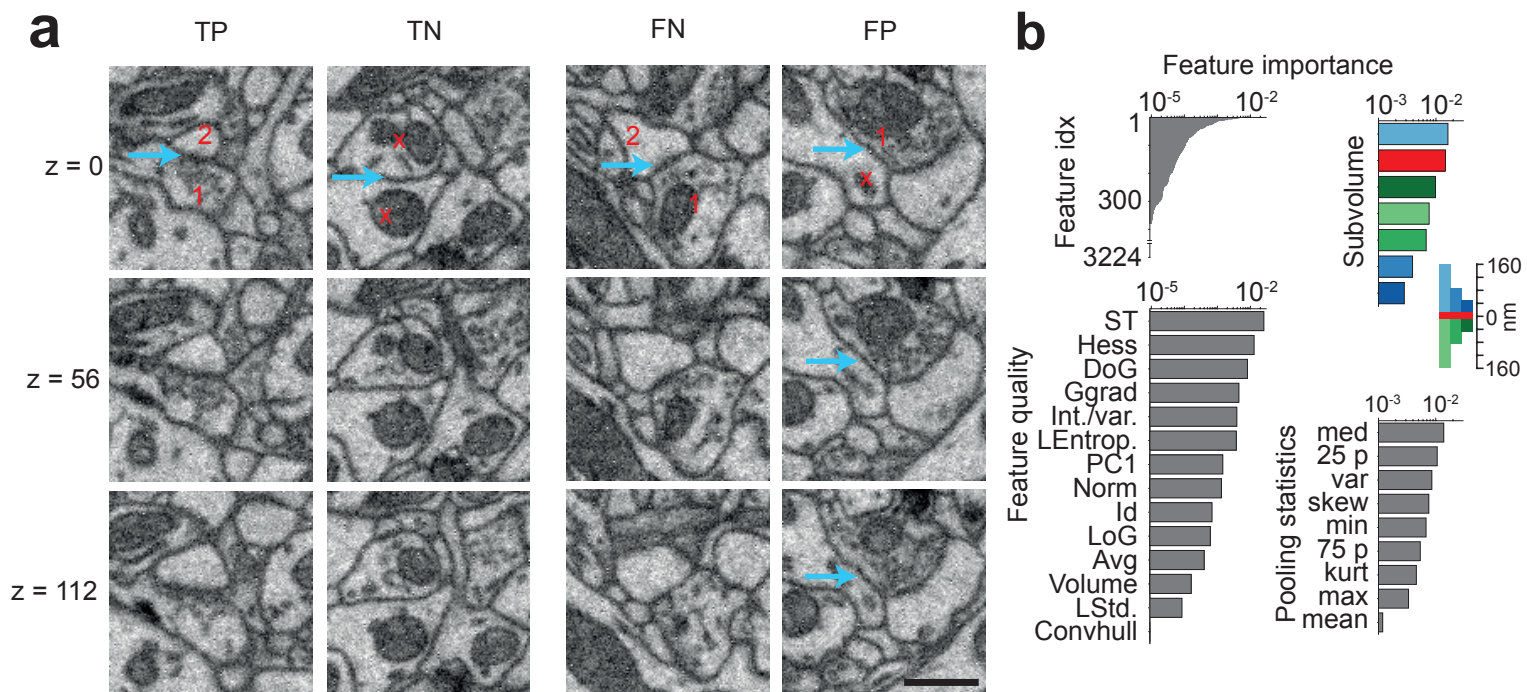


Figure 4  
Staffler et al.

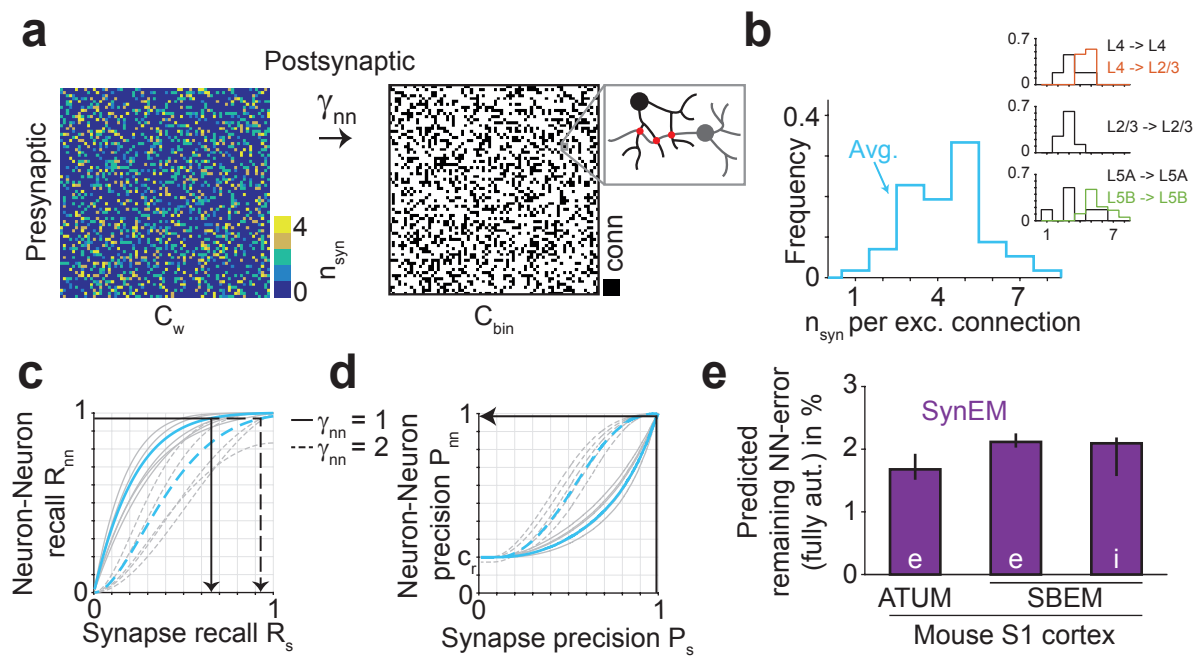


Figure 5  
Staffler et al.

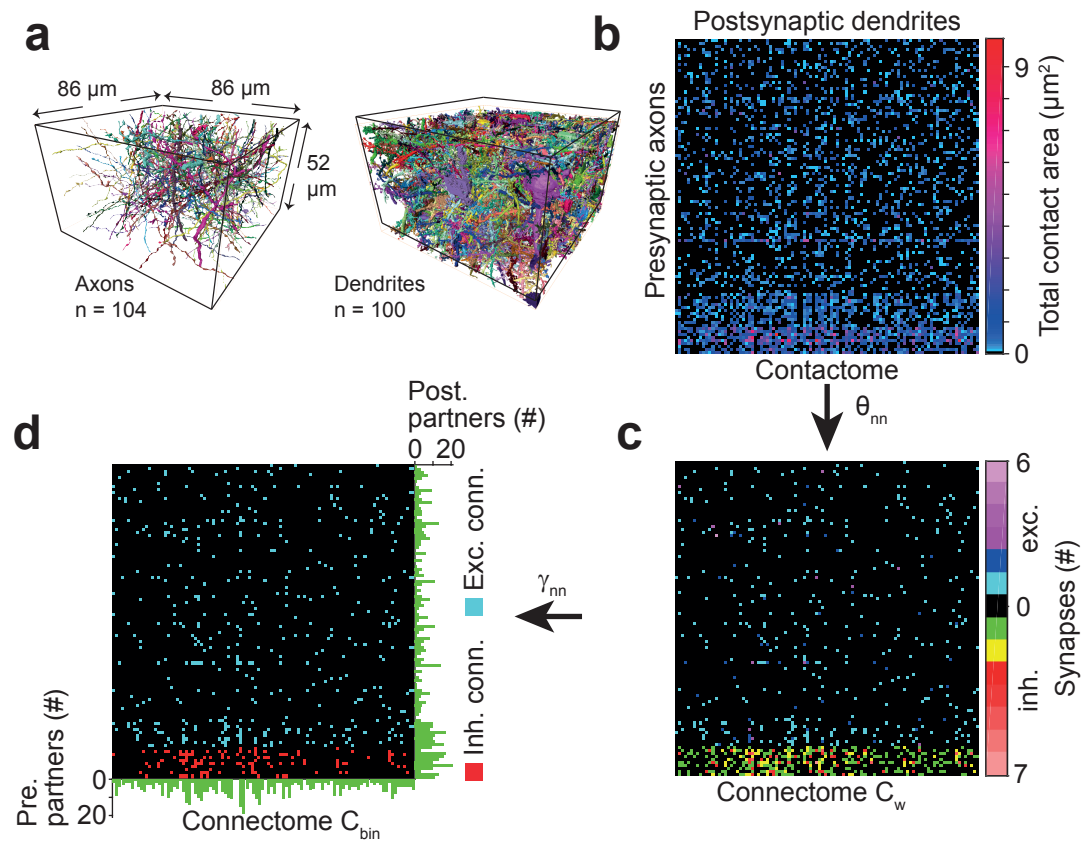


Figure 6  
Staffler et al.

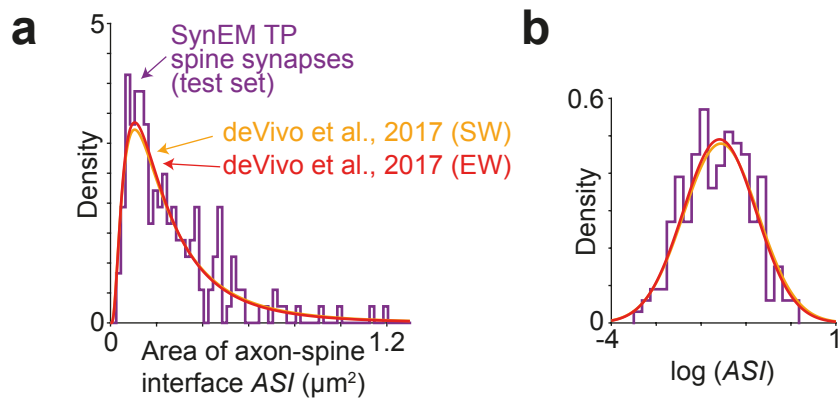


Figure 7  
Staffler et al.

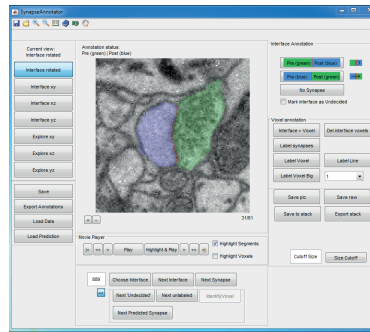


Figure 3 - figure supplement 1  
Staffler et al.

Graphical user interface (implemented in MATLAB) for efficient annotation of neurite interfaces as used for generating the training and validation labels. 3D image data is centered to the neurite interface and rotated such that the second and third principal components of the neurite interface span the displayed image plane. Segments are indicated by transparent overlay (interface, red; subsegment S1, blue and S2, green). Note that the test labels were independently annotated by volume search by multiple experts in webKnossos (Boergens et al., 2017), see Methods.

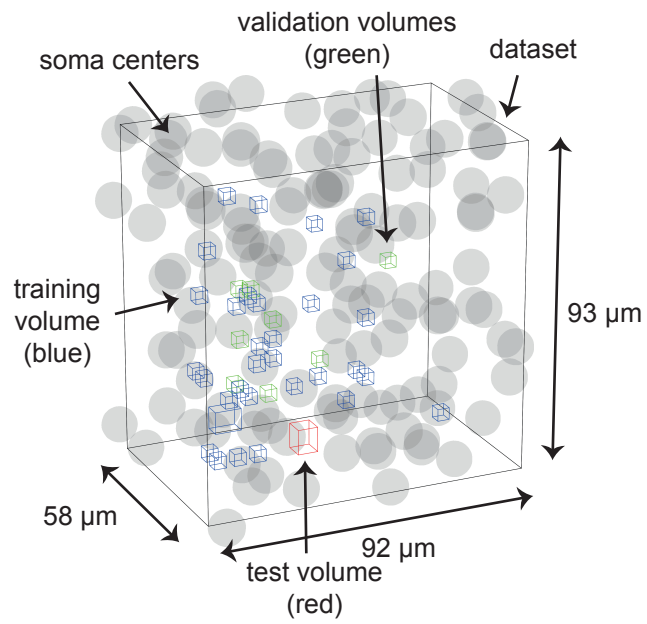
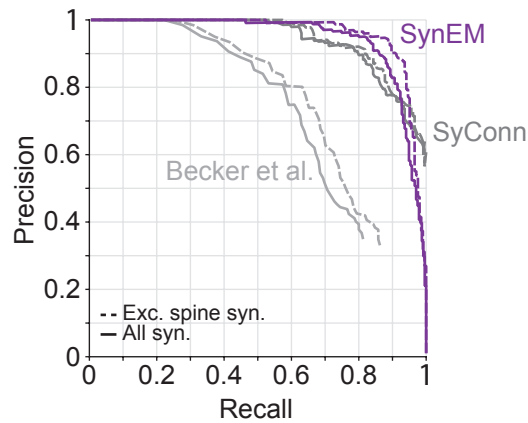


Figure 3 - figure supplement 2  
Staffler et al.

Distribution of training, validation and test data volumes within the dataset ex145\_07x2. Soma locations are indicated by spheres of radius 5  $\mu\text{m}$ .



**Figure 3 - figure supplement 3**  
**Staffler et al.**

Synapse detection performance comparison of SynEM with SyConn (Dorkenwald et al., 2017) and (Becker et al., 2012) on the 3D SBEM SynEM test set (Figure 3e). Note that while SynEM performs synapse detection and partner detection in one step these are separate steps in SyConn with an overall performance that is potentially different from the synapse detection step (in Dorkenwald et al., 2017, a reduction in performance by 10% in recall and 3% in precision from synapse detection to partner detection is reported, yielding a drop in F1 score of 0.057). Becker et al., 2012, does not contain a dedicated partner detection step.

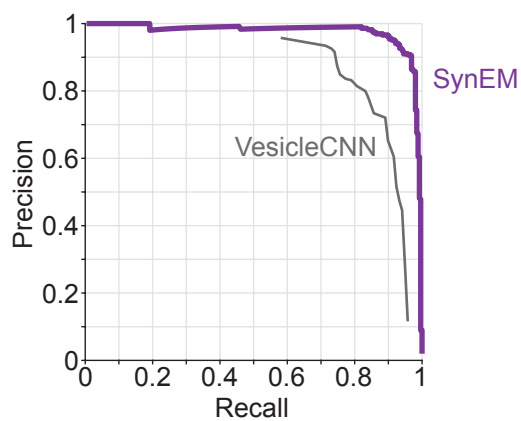


Figure 3 - figure supplement 4  
Staffler et al.

Synapse detection performance comparison of SynEM with VesicleCNN (Roncal et al., 2015) on a 3D EM dataset from mouse S1 cortex obtained using ATUM-SEM (Kasthuri et al., 2015). Note that VesicleCNN was developed on that ATUM-SEM dataset.



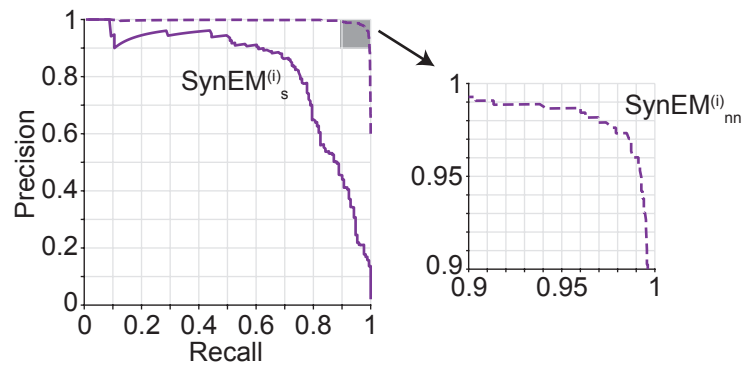
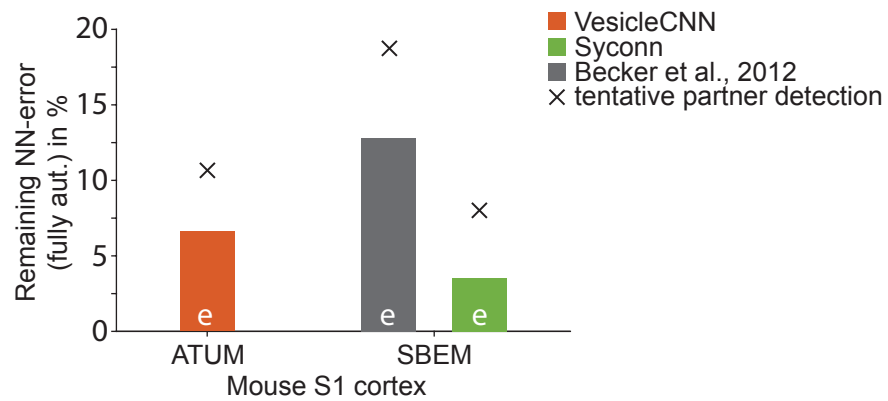


Figure 5 - figure supplement 1  
Staffler et al.

Performance of SynEM on a test set containing all interfaces between 3 inhibitory axons and all touching neurites (total of 9430 interfaces, 171 synapses). Single synapse detection precision and recall (solid line) and the ensuing predicted neuron-to-neuron precision and recall for inhibitory connections (dashed line) assuming on average 6 synapses for connections from interneurons (see Methods).



**Figure 5 - figure supplement 2**  
**Staffler et al.**

Effect of synapse detection errors on predicted connectome error rates for competing methods. Predicted neuron-to-neuron errors (reported as  $(1 - F1 \text{ score})$  in percent) for the ATUM-SEM dataset (Kasthuri et al., 2015) using VesicleCNN (Roncal et al., 2015, orange) and for our SBEM dataset using Becker et al., 2012 (gray) and Syconn (Dorkenwald et al., 2017, green). Note that these approaches provide synapse detection, only. When including the detection of the synaptic partners, Dorkenwald et al., 2017 reported a drop of detection performance by 3% precision and 10% recall (indicated by gray crosses, tentatively also for the other approaches). SynEM provides synapse detection and partner detection together (compare to Fig. 5e).

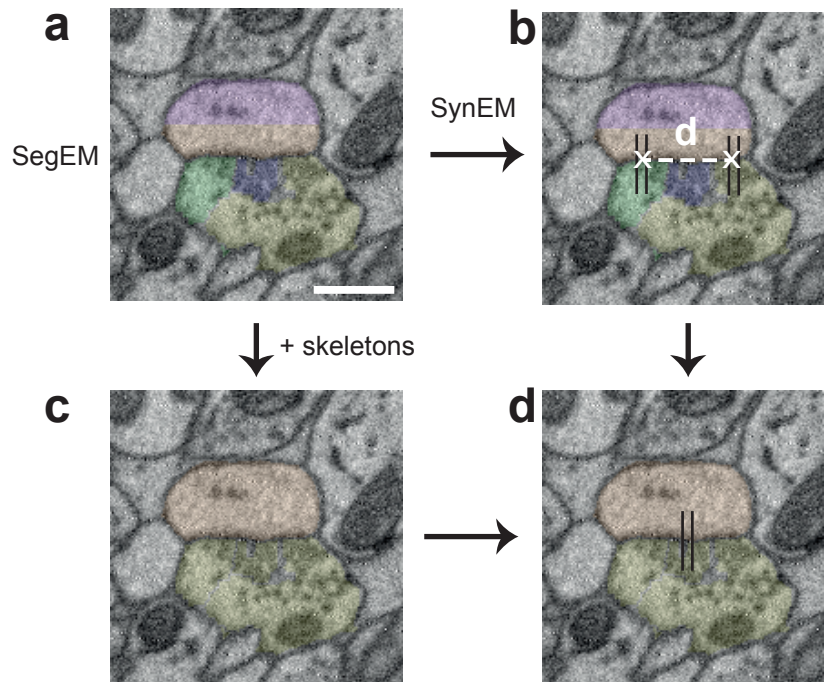


Figure 6 - figure supplement 1  
Staffler et al.

Procedure for obtaining synapse counts in the local connectome (Fig. 6). (a) Segmentation used for SynEM (note that a segmentation biased to neurite splits was used, see Berning et al., 2015) and (b) interfaces detected as synaptic (black lines). (c) combined skeleton-SegEM segmentation of neurites. (d) Synaptic neurite interfaces established between the same pre- and postsynaptic processes (as determined by the skeleton-SegEM segmentation, c) were clustered using hierarchical clustering with a distance cutoff of  $d = 1.5 \mu\text{m}$  (b) for obtaining the final synapse count. Scale bar, 500 nm.

## 7 Summary & Discussion

Understanding the brain and its unique abilities is a major preoccupation of many scientists today. This broad effort includes many disciplines like biology, computer science, physics, psychology, chemistry, mathematics and medicine. For this endeavor, mapping the synaptic circuits of any given brain region densely at cellular resolution could provide a valuable additional constraint to select between or rule out certain models of how brains work. Currently the main challenge to achieve this goal is to extract the synaptic connectivity graph from a given piece of brain tissue. The results in this thesis present methodological developments in the reconstruction of such synaptic circuits from 3D electron microscopy data.

This involves at least two important parts, one is neurite reconstruction and the other one is detecting the chemical synapses between them. The method for wire reconstruction presented here is split up in an automated part, used to generate locally accurate volume segmentations, and a manual part, for generating sparse annotations that provide the necessary long range information to join these local volumes. [Berning et al., 2015] focuses on the automated part while [Boergens et al., 2017] focuses on speeding up the manual annotation. The task of synapse detection on the other hand was investigated by phrasing the problem as a classification task for borders between segments [Staffler et al., 2017] which yielded sufficient performance to generate a binary connectome with 97% precision and recall without manual interaction.

### 7.1 From sparse annotations to volume models

#### Summary

[Berning et al., 2015] presents a semi-automated reconstruction workflow to generate volume models of cells in conventionally stained 3D electron microscopy data. It provides example reconstructions of synaptic networks in fully stained datasets from mouse retina and cortex and gives a detailed explanation of the workflow. The

reconstruction method presented is approximately 10-times faster than other methods published on conventionally stained data. Furthermore, a method was developed and tested to perform proximity detection on the reconstructed neurons to guide manual synapse validation to all close appositions of the reconstructed cells. Synapse validation was not possible using the cell-membrane only staining from previous studies [Briggman et al., 2011] [Helmstaedter et al., 2013]. These developments in combination resolved the trade-off between direct synapse validation and fast annotation of volume models in fully stained 3D electron microscopy data.

The automated part of the volume reconstruction approach involved a custom written CNN to detect membranes in fully stained SBEM data and generating a segmentation using marker based watershed on the results of the membrane CNN. This provided an over-segmentation of the neurites in the dataset, which could then be used in conjunction with sparse manual annotations providing longer-range information of neurite identity to create volume models. The sparse manual annotations annotate the center line of each branch of the neuron with points set at arbitrary distances resulting in a representation of a connected graph of nodes embedded in a 3D space termed skeleton reconstruction. The hyper-parameter optimization strategy for training the CNN as well as for optimizing the parameters in the marker generation for the watershed segmentation is described. These parameter selections were based on the split-merger metric described in the following paragraph.

The split-merger error metric was used to optimize the results of the automated procedures of membrane detection and subsequent segmentation on a training set and was afterwards measured on a test set. The metric uses the overlap matrix between manual skeleton annotations of all neurites in the dataset and an automated segmentation to calculate two types of errors: A split error was defined as each occurrence of a manually annotated skeleton overlapping with more than one segmentation volume object, while a merger error is defined as each instance of a segmentation object traversed by more than one manual skeleton annotation. The results were additionally evaluated and compared using the metrics of pixel, rand and warping error as proposed in the ISBI2012 challenge.

All data, especially the manually generated dense ground truth segmentations used for training the CNN and the skeleton annotations used for hyper-parameter optimization and evaluation of the segmentation were provided with the paper. Synaptic circuits that could be mapped with this approach were presented and details about the timelines in imaging, computation and manual analysis were discussed.

## Discussion

A similar reconstruction technique as described in [Berning et al., 2015] for generating volume models from sparse annotations had been used in [Helmstaedter et al., 2013], but in cell-membrane-only stained 3D electron microscopy data from mouse retina which lacked the benefit of direct synapse identification. The approaches presented [Berning et al., 2015] were based on the results from [Helmstaedter et al., 2013] and CNN architectures [Turaga et al., 2010] but extended their applicability to other staining methods and brain regions and provided the complete toolset needed for mapping synaptic circuits in such datasets.

The affinity maps introduced in [Turaga et al., 2010] were implemented and tested on the retina dataset in [Berning et al., 2015]. Affinity maps use a probability of each neighboring pair of voxel along the three cardinal axes to belong to the same neurite as an alternative representation for the intermediate result of image classification. We also classified the voxel directly into membrane vs. non-membrane voxel for the cortex dataset in [Berning et al., 2015]. Earlier implementations of CNN used for handwritten digit recognition [LeCun et al., 1989] were used as guides for weight initialization, choice of non-linearity and other model parameters.

Multiple improvements, alternative approaches or extensions to the automated methods for generation of an over-segmentation of neurites [Berning et al., 2015] have since been published. Just as [Berning et al., 2015] demonstrated the applicability of this method to fully stained EM datasets, [Pallotto et al., 2015] showed that extracellular space preservation can improve automated segmentation performance as the interface area between cells decreases and mergers can be more easily avoided.

Other improvements for partial automation of dense reconstruction are based on improvement of the segmentation procedure itself including the intermediate image classification step. Most notably Flood Filling Networks [Januszewski et al., 2016] aim to combine the two steps of membrane detection and watershed segmentation in a single step using a recursive 3D CNN. These models provide the advantage that end-to-end learning of image segmentation is possible without the intermediate step, while their current disadvantage is the higher computational cost. An improvement in error rates for segmentation was reported for a new model presented in [Funke et al., 2017], which extends the approach presented in [Turaga et al., 2009] and uses the training data and dense skeleton reconstructions published in [Berning et al., 2015] as one of three test cases for the method.

Note that [Januszewski et al., 2016] and [Funke et al., 2017] also suggest

an improvement and an alternative to the split-merger metric described in [Berning et al., 2015]. The alternative metric had previously been described in [Funke et al., 2015] and the authors neither discuss nor compare to a node threshold of 1 as introduced in [Berning et al., 2015] when dismissing the node threshold of 2 as too high for their performance measurements [Funke et al., 2017] and rather use their previously published method. The improvement to the split-merger metric is presented with a strong argument [Januszewski et al., 2016] and will thus probably yield a better metric for segmentation performance. Note that this argument is only important in the very under-segmented regime, which means that many merge errors are present. This regime is uninteresting in the biological sense as it would not provide any help with the reconstruction effort and all segmentations presented in [Berning et al., 2015] were rather biased into the regime of over-segmentation as discussed in detail in the paper. A rigorous comparison between all alternative metrics has not yet been performed.

The analysis of reconstruction speed using sparse annotations reported in [Berning et al., 2015], has since been used to estimate reconstruction effort needed for whole mouse brain reconstructions [Mikula, 2016]. Furthermore the training data published with this work have since been used to train adversarial networks that aim to optimize the alignment of the dataset [Jain, 2017], during which the images taken by the microscope are put into a 3D reference frame. This is usually done by means of cross-correlation or matching of image features. In the case of the cortex dataset in [Berning et al., 2015] a modified version of [Preibisch et al., 2009] was used. As alignment is of crucial importance for all further 3D processing of the data, it is likely that this improvement would benefit all subsequent processing steps.

In summary, while improvements of the error rate in the automated segmentation presented in [Berning et al., 2015] as well as improvements or alternatives to the split-merger metric have been published, no new method for the extraction of neural wires based on these automated results have been published since. This would suggest that while one might choose a more accurate automated segmentation technique, the rest of the workflow presented in the paper should still be considered the state of the art for the generation of volume models and synaptic circuits from 3D EM data.

## 7.2 Novel annotation interactions

### Summary

[Boergens et al., 2017] presents an in-browser annotation tool for viewing and annotating 3D electron microscopic data. The methodological improvement over the existing in-browser annotation tool CATMAID [Schneider-Mizell et al., 2016] was highlighted and the importance of efficient streaming of electron microscopy data for crowd-sourcing was discussed. The main result of these methodological improvements is enabling seamless navigation through the data with minimal bandwidth and latency requirements. Together with a novel egocentric single view-port reconstruction method, "flight mode", this tool enables trained annotators to create skeleton annotations at a speed of  $2.1 \pm 0.9$  mm/h (mean  $\pm$  std) for dendrite trunks and  $1.5 \pm 0.6$  mm/h (mean  $\pm$  std) for axons in a SBEM dataset from mammalian cerebral cortex. This speeds up skeleton annotation by a factor of 4-13 over published methods. Extensive quantification of the error rates and annotation speed in two annotation modes were performed, and comparisons to speeds and error rates published with other reconstruction tools were made. Furthermore a workflow for extraction of a synaptic circuit from 3D EM data is outlined and applied to a SBEM dataset.

The need for online delivery of data for crowd sourcing of data analysis is evident when considering the size of currently published 3D EM datasets with sizes of up to 100 terabytes [Zheng et al., 2017] and accepting the premise that continuous annotation of whole cells should be possible for each annotator. This would otherwise require tens of hard-drives to be delivered to each annotator. In order to minimize bandwidth and latency requirements webKnossos loads data close to the current position of the annotator first. This is enabled by a new data storage and transmission format in small  $(32\text{voxel})^3$  cubes refereed to as "buckets". Any 2D data storage and transmission by design prioritizes the data locality along a 2D plane within the 3D volume and will thus prioritize loading data further away from the current point of interest along a given dimension and thus require longer time to display the needed information.

The transmission and local assembly of these buckets into one reference frame allows for arbitrary views in this data. This fact was used to develop a novel reconstruction mode, called "flight mode". In this mode the user is virtually placed in a hollow sphere carved out of the dataset and can freely choose the current orientation of viewing and move the sphere forward or backward throughout the dataset along the current viewing orientation. This egocentric view mode seemed much more intuitive than the 3



axis-aligned view-ports usually used called "ortho mode". One can try both annotation approaches at <https://webknossos.org>.

Comparisons of tracing speed and accuracy of flight and ortho mode revealed that this new annotation mode enables faster tracing at indistinguishable error rates. In order to quantify error rates at different tracing redundancies, the redundant skeleton consensus procedure (RESCOP) algorithm [Helmstaedter et al., 2011] was used to consolidate multiple redundant annotations of the same process. This shows that even trained annotators have an error rate of 14 errors/mm at single redundancy when counting all errors and about 7 errors/mm when only counting errors due to which more than 10 micron path length is lost or added when tracing axons. This error number can then be scaled down by increasing the redundancy in the annotation which shows that most errors are uncorrelated, see also [Helmstaedter et al., 2011] for an earlier quantification of tracing accuracy on a different dataset and an introduction of the RESCOP algorithm.

The connectome reconstruction workflow proposed in [Boergens et al., 2017] does not require any automated processing, and is thus the recommended way of analysis for datasets of quality not deemed sufficient for automation or if initial investments in automation is not justified by the reconstruction goal. Note that the annotation tool can also be used for semi-automated annotation approaches, but this is not emphasized in the paper for reasons of focusing on the core improvements. The first step for connectome reconstruction is to trace all processes in the dataset at a given redundancy, for example in the paper redundancies of 3 and 6 are used for dendritic trunks and axons respectively. The redundancy should be chosen according to the error tolerance for the biological question under investigation. The redundant tracings are then consolidated using the RESCOP algorithm. Afterwards all synapses made by the consolidated axons were annotated manually by clicking the postsynaptic process in a viewing mode that automatically follows the consolidated skeleton. These locations are then used as seed locations for tracing the postsynaptic process, usually a spine head, back to the dendritic trunk, which is then automatically matched to the dendrite trunks annotated before. This provides all information required to assemble a connectome. In this paper all code is provided to set up a server deploying this tool and all auxiliary code to perform the connectome assembly, path length, speed and error measurements.

The main result is the speedup of human data annotation in browser for EM based connectomics by 4-13 fold over published results and that these theoretical speed gains can be achieved by a human annotators. The error quantification additionally shows that this is possible without a loss in tracing accuracy. The connectome workflow shows

how to use this method to reconstruct synaptic circuits within an EM dataset.

## Discussion

The improvements in annotation speed coupled with the browser based data delivery have enabled projects that would not have been feasible using the old reconstruction methods or means of data distribution. Most prominent is the recent discovery of a sorting of output synapses along the proximal part of axons of excitatory cells in medial entorhinal cortex, which first target inhibitory cells and later other excitatory cells [Schmidt et al., 2017]. The paper furthermore describes that these targeted inhibitory cells in turn make synapses back onto the excitatory population suggesting a very concrete circuit implementation of cellular feed-forward inhibition in cortex, including wide diameter myelinated inhibitory axons as a biophysical mechanism for fast transmission of this feed forward inhibition. All data annotation for this project was performed using webKnossos [Boergens et al., 2017].

A total path length of 2.89 m dendritic trunks was traced in 3,654 work hours in the orthogonal tracing mode [Schmidt et al., 2017]. Note that this yields a tracing speed of 0.79 mm/h, slightly lower than for the speed measurements performed in [Boergens et al., 2017]. As also discussed in the latter paper this could be attributed to either the larger dataset with worse image and alignment quality or motivation of the human annotators over long annotation projects in general. Also note the measured error rates at 2-fold redundancy reported in [Boergens et al., 2017] would suggest that the tracings of dendrites at "average redundancy of 2.0" [Schmidt et al., 2017] contain an average of approximately 3 errors per mm traced path length. [Schmidt et al., 2017] show that the efficient online data delivery presented in [Boergens et al., 2017] enabled the collaboration of more than 20 people on a dataset of approximately 40 terabyte and provides valuable numbers about possible speed gains and error rate considerations for such reconstruction projects.

The approach for connectome generation presented in [Boergens et al., 2017] does not require any automation and can thus be used for fast sparse annotations of synaptic networks as in [Schmidt et al., 2017]. If either volume models of neurons [Berning et al., 2015] or automated synapse detection [Staffler et al., 2017] is needed, webKnossos provides additional features to visualize the results generated with these methods. In general the decision for or against partial automation of the analysis depends on the quality of the dataset, because worse alignment or image quality will generate higher error rates in the automated part while human annotators can usually

ignore these to a some extent, and on the overall expected manual annotation workload as automation involves some overhead like training data generation and deployment and possible modification of algorithms.

It is very important to note that [Boergens et al., 2017] mainly focuses on the reconstruction of axons. This is due to the fact that the error rates in dendrites are much lower and dendritic trunk path length is only a fraction of axon path length. Axon reconstruction thus presents the main challenge in the generation of cortical connectomes in mammals. Also note that the most difficult part in reconstruction of dendrites is the tracing of the spine necks which can get very thin [Helmstaedter, 2013].

The methods described in [Boergens et al., 2017] have only been tested on a single SBEM datasets and it is not yet clear which quality of alignment is needed for the novel annotation interactions. The new annotation mode "flight mode" is likely to be useful in any context where the imaged volume can be sufficiently well aligned to guarantee 3D membrane continuity. Reconstructions in datasets, which do not meet these requirements, can still profit from the speed improvements in the "ortho mode" annotation approach.

## 7.3 Automated synapse detection

### Summary

[Staffler et al., 2017] presents a method for automatic inference of synaptic locations and directionality in 3D EM data. The classifier achieved a performance of 88% precision and recall on single synapses. Furthermore, a model is introduced which shows that the method will achieve 97% recall and precision in a binary connectome when taking into account synapse count distributions between pairs of cells from the literature.

The method is based on classifying interfaces, the surface between two neighboring segmentation objects, as either synaptic or non-synaptic. The method thus requires a segmentation, in this case one of those presented in [Berning et al., 2015] was used for all experiments, except for the experiments described on the ATUM dataset from [Kasthuri et al., 2015], where a manual segmentation provided with the dataset was used. A representation of these interfaces is created by defining different sub-volumes based on distance to the interface. This representation is used to calculate different summary statistics, moments of the distribution, as well as some rank-based measures on the pixel values of different 3D-filter applied to the raw data over these sub-volumes.

The paper describes in detail how the classifier was developed and continually improved by adding additional 3D-filter, sub-volumes and summary statistics. In addition to these extensions, slightly different ways of defining the classification tasks are described and it is shown that best performance can be achieved when using the intrinsic directionality of the synapse and classifying segment pairs in both directions and only predicting a synapse when the ordering of the segment pairs is from pre- to postsynaptic. Additional analysis of remaining error locations of the synapse detection and the importance of different sub-volumes, 3D-filter and summary statistics for the classification decision is provided.

The model that predicts the performance for a binary connectome based on classification performance on single synapses is motivated and the underlying literature is discussed. This shows that this method can be used for fully automated neuron to neuron connectivity inference with 97% precision and recall.

## Discussion

Synapse detection performance as presented in this paper is dependent on the dataset analyzed, especially on the staining and imaging method used as these take different trade-offs between imaging speed, emphasized intracellular structures and quality of the dataset. Focused ion-beam based SEM at an isotropic voxel size of 4-8nm makes automated synapse detection most feasible and low error rates have been reported previously [Kreshuk et al., 2011]. As this microscopy method can currently only be used to image small volumes, only trading off resolution for more imaged volume per time makes it plausible to start imaging whole neurons in the mammalian brain. This trade-off is for example encountered in ssTEM or ATUM based SEM, which sacrifice resolution mostly along one dimension, and diamond knife based SEM, which is usually used with a more isotropic trade-off on resolution in all dimensions.

The approach presented in [Staffler et al., 2017] outperforms other approaches for automated synapse detection on a SBEM dataset from mouse cortex, most notably the previous state of the art models described in [Dorkenwald et al., 2017] and [Roncal et al., 2014]. The transferability of the method and its good performance are shown on a ATUM based SEM [Kasthuri et al., 2015] dataset. Note that transferring the method between these two datasets requires retraining the classifier, which is to be expected for datasets from such different imaging modalities. This is an indication that a segmentation based interface classification as proposed in [Staffler et al., 2017] can provide superior performance than methods which classify single voxel, for example

[Dorkenwald et al., 2017].

Because the approach predicts a pre- and postsynaptic segment it intrinsically also provides synaptic partner detection, e.g. predicting a region that belongs to the pre- and postsynaptic process respectively, which has to be incorporated in a different step when only classifying single voxel as synaptic and can lead to a drop in overall performance, as documented in [Dorkenwald et al., 2017].

As for wire reconstruction, automated synapse detection performance can be expected to improve if extracellular space preservation [Pallotto et al., 2015] is used because touch of two processes will provide a stronger indication for a synapse. Furthermore the surface for false positive detection is decreased. The method presented here is itself sufficient for generating binary connectomes in conventionally stained data.

It is important to note that so far the analysis of single electron micrographs, usually imaged using a TEM, have been regarded as the gold standard for establishing the existence of a synapse in a certain location. This is usually based on the darkening of the plasma membrane close to the surface in the postsynaptic process, the postsynaptic density, and vesicles close to or even docked to the plasma membrane in the presynaptic process. The advantage of using a high-resolution single plane electron micrograph for this purpose is that these intracellular structures are more easily resolved than using a 3D image stack with less resolution along the imaging dimensions, especially when taking into account the volume vs. resolution tradeoff taken in datasets aimed at the reconstruction of cortical circuits in mammals discussed above. The disadvantage of the methods with high in-plane resolution is their worse resolution along the cutting direction [Briggman and Bock, 2012]. This will make the detection of synaptic interfaces parallel to the imaging plane more difficult.

The main advantage of 3D image microscopy methods is that a lot of contextual information along all dimensions can be used for judging whether a certain location is synaptic. For this reason [Staffler et al., 2017] provides a comprehensive guide for identifying synaptic locations in SBEM data in the supplementary material. In general, the quantification of synapse detectability in different 3D-EM datasets is still outstanding, and all precision and recall measurements reported in [Staffler et al., 2017] are based on the the annotation of synapses by expert annotators using the criteria presented in the supplementary material.

## 7.4 General discussion

While I have presented my view of why connectomics is and will be important for neuroscience in the introduction, and presented methods for easier generation of connectomes based on electron microscopy in the results, it is also important to discuss the limitations of the approach of dense EM reconstructions. Note that due to renewed interest for these topics in the recent decade, these limitations are likely to shift with further method development in image acquisition and analysis.

Currently the most important consideration in all electron microscopy approaches is the tradeoff between volume that can be imaged versus resolution and image quality. Furthermore, each imaging technique has different inherent image artifacts which can lead to additional challenges in data analysis or even make dense mapping of circuits impossible. The methods summarized in this thesis have all been evaluated on SBEM data from mouse cerebral cortex. In [Berning et al., 2015] an additional mouse retina SBEM dataset was used for validation of the transferability of the method while in [Staffler et al., 2017] an additional ATUM dataset originally presented in [Kasthuri et al., 2015] was used for for this purpose. Both of these methods for acquiring 3D EM data are aimed at acquiring volumes at a resolution just sufficient for dense reconstruction and are thus able to image larger volumes than using focused ion-beam based SEM. The SBEM datasets presented in [Berning et al., 2015] have a volume of approximately 0.5 million  $\mu m^3$ , while the ATUM dataset in [Kasthuri et al., 2015] spans about 80.000  $\mu m^3$  in highest resolution. Note that larger datasets have been published, but rigorous quantification of reconstructability across these datasets is still missing.

No matter which 3D electron microscopy technique is used to acquire data, an inherent tradeoff between imaged volume, resolution and quality of the dataset, e.g. with respect to signal-to-noise ratio of single voxel or imaging artifacts, can be observed. For the non block-face methods this is particularly pronounced for the cutting dimensions as some sections are usually lost or damaged during the slicing procedure and all datasets with a volume aimed at circuit reconstruction published so far have even reported multiple consecutive slice losses [Bock et al., 2011] [Kasthuri et al., 2015] [Lee et al., 2016]. This tradeoff can also be noted for other methods when comparing e.g. [Hildebrand et al., 2017] or [Schmidt et al., 2017] to the smaller datasets of the respective method as presented in [Kasthuri et al., 2015] or [Boergens et al., 2017].

The volume of current datasets clearly limit the insights that can be gained. For example, no known neuron in mammalian cerebral cortex has dendritic or axonal

arborisations restricted to such a small volume as for the datasets presented in this thesis. Therefore, current connectomic studies in mammalian cortex are limited by slicing artifacts and can only answer a small subset of questions about connectivity. Some recent methodological improvements for image acquisition like mSEM [Eberle et al., 2015], development of high throughput ssTEM [Zheng et al., 2017] or the hot knife technique [Hayworth et al., 2015] show approaches to overcome the volume limitations of some image acquisition techniques. So far, none of these approaches have yielded datasets of a size large enough to encompass whole neurons with sufficient data quality for even manual dense annotation. Current automated methods require even better quality. For non block-face approaches and the hot knife technique it further remains unclear how well the continuation of neurites over lost, damaged and folded slices or hot knife cuts can be established. See [Hayworth et al., 2015] for first quantifications of this issue for the hot-knife technique.

Furthermore, EM based connectomics will only ever be able to yield a snapshot of the connectivity in any given animal. This is due to the nature of the experimental technique of sample preparation, which fixes the tissue block at a given time point. This is especially challenging as it has not yet been quantified how reproducible the synaptic connectivity graph is or which properties of it might (not) be conserved. It has been observed that the lifetimes of spines vary greatly with only about 50% being present over the imaging period of a month [Trachtenberg et al., 2002]. Therefore a quantification of the inter-individual variability in the connectivity graph and a study of its conserved properties is of great importance to the field of EM based connectomics.

Finally, while methods exist to image a given volume using a fluorescence microscope prior to connectomic circuit reconstruction and map their functional responses to stimuli [Briggman et al., 2011], one should note that most connectomic datasets do not provide that information and additional challenges in the sample preparation are usually a consequence. It has not yet been shown that these correlated datasets do not generate a trade-off between functional data and a good preservation of the ultrastructure which is essential for (partial) automation of the reconstruction.

## 7.5 Outlook

Especially after the improvements in automated synapse detection [Staffler et al., 2017] presented in the last section of this thesis, the major bottleneck for generating connectomes from 3D EM data is generating volume models of the neurites which

provide the necessary information to connect these synapses. The approach of manual sparse skeleton annotations in combination with a locally accurate segmentation [Berning et al., 2015] still requires a human annotator to follow the neurite throughout the dataset.

While the speed of this manual annotation was improved significantly [Boergens et al., 2017], it still limits the path length that can be achieved during a given manual reconstruction time and even reconstructing the approximately 4 kilometers of axons and 1 kilometer of dendrites [Braitenberg and Schüz, 2013] in a cubic millimeter of cortical tissue at one fold redundancy would require more than one million hours manual annotation effort. When considering that higher redundancy might be required, we can conclude that further method development in data analysis for wire reconstruction from 3D-EM data is required.

One approach to further improve reconstruction speed is to focus human annotation on locations that the current algorithms have problems in solving automatically and to use this information to gradually improve the algorithms. This in turn will focus human attention on more difficult locations over time and provide further gains in reconstruction speed by eliminating the need for a human annotator to inspect each location of a neurite manually.

Therefore, developing a method for "focused annotation" of neurites has been the focus of my work recently. The approach is based on the classification on the level of interfaces, as in [Staffler et al., 2017], but with the question whether the local interface between two segments should be a continuation of the neurite. This could be called a "neurite continuity classifier" for the segment graph. Each pair of segments that is merged based on such a classifier would remove one split error in the metric presented above and would not have to be annotated manually. In addition to this neurite continuity classifier, a method is required to determine the location of the two types of errors, splits and mergers, which the initial over-segmentation might introduce [Berning et al., 2015]. For this an algorithm which detects endings in cylindrical elongated structures was developed and is currently evaluated. This information is then used to focus human annotation to these positions to fix split errors by manual annotation. Furthermore, some notion of a merger detector is required as well. For this purpose we currently use a detector of all intersections in neurite morphology with more than 4 exits, which would suggest a merger error. As the biological frequency of such configurations is much lower than current merger rates, the additional annotation of the error free locations is not yet an issue.



While I think that complete automation of the analysis process is desirable, I do not think that this likely in the foreseeable future, especially for datasets with numerous artifacts and varying quality as currently present in datasets of sufficient volume for circuit reconstruction in mammalian cortex. I therefore believe that the automated guidance of manual annotation to difficult locations or errors in a current state of the automation on the dataset will be the main approach for connectomic data analysis in the foreseeable future and I have presented approaches here which could be extended in this way.

The field of connectomics in general will have to validate some of its approaches. The reconstructability of wires in a given dataset, as well as the detectability of synapses needs to be formalized in some way to allow for calculation and discussion of these quantities for each biological question and comparison of the accuracy of reconstruction between different studies in a reliable manner. Furthermore the inter-individual variability between animals needs to be quantified and quantities derived which are preserved for animals that have the same genetics and were raised in the same environment even if their overall connectivity graph varies.

## 8 Bibliography

- [Ahrens et al., 2013] Ahrens, M. B., Orger, M. B., Robson, D. N., Li, J. M., and Keller, P. J. (2013). Whole-brain functional imaging at cellular resolution using light-sheet microscopy. *Nature methods*, 10(5):413–420.
- [Ammer et al., 2015] Ammer, G., Leonhardt, A., Bahl, A., Dickson, B. J., and Borst, A. (2015). Functional specialization of neural input elements to the drosophila on motion detector. *Current Biology*, 25(17):2247–2253.
- [Arkley et al., 2014] Arkley, K., Grant, R. A., Mitchinson, B., and Prescott, T. J. (2014). Strategy change in vibrissal active sensing during rat locomotion. *Current Biology*, 24(13):1507–1512.
- [Azevedo et al., 2009] Azevedo, F. A., Carvalho, L. R., Grinberg, L. T., Farfel, J. M., Ferretti, R. E., Leite, R. E., Lent, R., Herculano-Houzel, S., et al. (2009). Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *Journal of Comparative Neurology*, 513(5):532–541.
- [Baden et al., 2013] Baden, T., Berens, P., Bethge, M., and Euler, T. (2013). Spikes in mammalian bipolar cells support temporal layering of the inner retina. *Current biology : CB*, 23:48–52.
- [Barth et al., 2016] Barth, A. et al. (2016). Comment on: Principles of connectivity among morphologically defined cell types in adult neocortex (vol 353, pg aaj2094, 2016). *Science*, 354(6318):1386–1386.
- [Behnia et al., 2014] Behnia, R., Clark, D. A., Carter, A. G., Clandinin, T. R., and Desplan, C. (2014). Processing properties of on and off pathways for drosophila motion detection. *Nature*, 512(7515):427.
- [Berning et al., 2015] Berning, M., Boergens, K. M., and Helmstaedter, M. (2015). SegEM: efficient image analysis for high-resolution connectomics. *Neuron*, 87(6):1193–1206.

- [Binzegger et al., 2004] Binzegger, T., Douglas, R. J., and Martin, K. A. (2004). A quantitative map of the circuit of cat primary visual cortex. *Journal of Neuroscience*, 24(39):8441–8453.
- [Bobrov et al., 2014] Bobrov, E., Wolfe, J., Rao, R. P., and Brecht, M. (2014). The representation of social facial touch in rat barrel cortex. *Current Biology*, 24(1):109–115.
- [Bock et al., 2011] Bock, D. D., Lee, W.-C. A., Kerlin, A. M., Andermann, M. L., Hood, G., Wetzel, A. W., Yurgenson, S., Soucy, E. R., Kim, H. S., and Reid, R. C. (2011). Network anatomy and in vivo physiology of visual cortical neurons. *Nature*, 471:177–182.
- [Boergens et al., 2017] Boergens, K. M., Berning, M., Bocklisch, T., Bräunlein, D., Drawitsch, F., Frohnhofen, J., Herold, T., Otto, P., Rzepka, N., Werkmeister, T., Werner, D., Wiese, G., Wissler, H., and Helmstaedter, M. (2017). webknossos: efficient online 3d data annotation for connectomics. *Nature Methods*, 14:691–694.
- [Borst and Helmstaedter, 2015] Borst, A. and Helmstaedter, M. (2015). Common circuit design in fly and mammalian motion vision. *Nature neuroscience*, 18:1067–1076.
- [Braitenberg and Schüz, 2013] Braitenberg, V. and Schüz, A. (2013). *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media.
- [Brecht et al., 2003] Brecht, M., Roth, A., and Sakmann, B. (2003). Dynamic receptive fields of reconstructed pyramidal cells in layers 3 and 2 of rat somatosensory barrel cortex. *The Journal of physiology*, 553(1):243–265.
- [Briggman and Bock, 2012] Briggman, K. L. and Bock, D. D. (2012). Volume electron microscopy for neuronal circuit reconstruction. *Current opinion in neurobiology*, 22:154–161.
- [Briggman et al., 2011] Briggman, K. L., Helmstaedter, M., and Denk, W. (2011). Wiring specificity in the direction-selectivity circuit of the retina. *Nature*, 471:183–188.
- [Broca, 1865] Broca (1865). Sur le siége de la faculté du langage articulé. *Bulletins de la Société d’anthropologie de Paris*, 6(1):377–393.
- [Cajal, 1888] Cajal, S. R. (1888). *Estructura de los centros nerviosos de las aves*. Jimenez y Molina.

- [Cardona et al., 2010] Cardona, A., Saalfeld, S., Preibisch, S., Schmid, B., Cheng, A., Pulokas, J., Tomancak, P., and Hartenstein, V. (2010). An integrated micro-and macroarchitectural analysis of the drosophila brain by computer-assisted serial section electron microscopy. *PLoS biology*, 8(10):e1000502.
- [Catania and Kaas, 1997] Catania, K. C. and Kaas, J. H. (1997). The mole nose instructs the brain. *Somatosensory & motor research*, 14:56–58.
- [Celikel and Sakmann, 2007] Celikel, T. and Sakmann, B. (2007). Sensory integration across space and in time for decision making in the somatosensory system of rodents. *Proceedings of the National Academy of Sciences*, 104(4):1395–1400.
- [Chen et al., 2015] Chen, F., Tillberg, P. W., and Boyden, E. S. (2015). Expansion microscopy. *Science*, 347(6221):543–548.
- [Cole and Curtis, 1939] Cole, K. S. and Curtis, H. J. (1939). Electric impedance of the squid giant axon during activity. *The Journal of general physiology*, 22(5):649–670.
- [Denk et al., 2012] Denk, W., Briggman, K. L., and Helmstaedter, M. (2012). Structural neurobiology: missing link to a mechanistic understanding of neural computation. *Nature reviews. Neuroscience*, 13:351–358.
- [Denk and Horstmann, 2004] Denk, W. and Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS biology*, 2:e329.
- [Denk et al., 1990] Denk, W., Strickler, J. H., Webb, W. W., et al. (1990). Two-photon laser scanning fluorescence microscopy. *Science*, 248(4951):73–76.
- [Diamond et al., 2008] Diamond, M. E., Von Heimendahl, M., Knutsen, P. M., Kleinfeld, D., and Ahissar, E. (2008). 'where'and'what'in the whisker sensorimotor system. *Nature Reviews Neuroscience*, 9(8):601–612.
- [Dörfl, 1985] Dörfl, J. (1985). The innervation of the mystacial region of the white mouse: A topographical study. *Journal of Anatomy*, 142:173.
- [Dorkenwald et al., 2017] Dorkenwald, S., Schubert, P. J., Killinger, M. F., Urban, G., Mikula, S., Svara, F., and Kornfeld, J. (2017). Automated synaptic connectivity inference for volume electron microscopy. *Nature Methods*, 14(4):435–442.

- [Doyle et al., 1998] Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T., and MacKinnon, R. (1998). The structure of the potassium channel: molecular basis of  $K^+$  conduction and selectivity. *Science (New York, N.Y.)*, 280:69–77.
- [Dreyfus, 1962] Dreyfus, S. (1962). The numerical solution of variational problems. *Journal of Mathematical Analysis and Applications*, 5(1):30–45.
- [Eberle et al., 2015] Eberle, A., Mikula, S., Schalek, R., Lichtman, J., Tate, M. K., and Zeidler, D. (2015). High-resolution, high-throughput imaging with a multibeam scanning electron microscope. *Journal of microscopy*, 259(2):114–120.
- [Eichler et al., 2017] Eichler, K., Litwin-Kumar, A., Li, F., Park, Y., Andrade, I., Schneider-Mizell, C. M., Saumweber, T., Huser, A., Eschbach, C., Gerber, B., et al. (2017). The complete connectome of a learning and memory center in an insect brain. *bioRxiv*, page 141762.
- [Euler et al., 2002] Euler, T., Detwiler, P. B., and Denk, W. (2002). Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature*, 418(6900):845.
- [Feldmeyer et al., 1999] Feldmeyer, D., Egger, V., Lübke, J., and Sakmann, B. (1999). Reliable synaptic connections between pairs of excitatory layer 4 neurones within a single barrelof developing rat somatosensory cortex. *The Journal of Physiology*, 521(1):169–190.
- [Flourens, 1824] Flourens, P. (1824). Experimental researches on the properties and functions of the nervous system in the vertebrate animal, 1824. *Appleton-Century-Crofts*.
- [Fox, 2008] Fox, K. (2008). *Barrel cortex*. Cambridge University Press.
- [Frégnac and Laurent, 2014] Frégnac, Y. and Laurent, G. (2014). Where is the brain in the human brain project? *Nature*, 513(7516):27.
- [Fritsch and Hitzig, 1863] Fritsch, G. and Hitzig, E. (1863). Electrical excitability of cerebrum. *Journal of Neurosurgery*, 20(10):905.
- [Fukushima, 1979] Fukushima, K. (1979). Neural network model for a mechanism of pattern recognition unaffected by shift in position- neocognitron. *Electron. & Commun. Japan*, 62(10):11–18.

- [Funke et al., 2015] Funke, J., Moreno-Noguer, F., Cardona, A., and Cook, M. (2015). Ted: A tolerant edit distance for segmentation evaluation. *arXiv*.
- [Funke et al., 2017] Funke, J., Tschopp, F. D., Grisaitis, W., Singh, C., Saalfeld, S., and Turaga, S. C. (2017). A deep structured learning approach towards automating connectome reconstruction from 3d electron micrographs. *arXiv*.
- [Galvani and Aldini, 1792] Galvani, L. and Aldini, G. (1792). *De Viribus Electricitatis In Motu Musculari Comentaribus Cum Joannis Aldini Dissertatione Et Notis; Accesserunt Epistolae ad animalis electricitatis theoriam pertinentes*. Apud Societatem Typographicam.
- [Garcia-Lopez et al., 2010] Garcia-Lopez, P., Garcia-Marin, V., and Freire, M. (2010). The histological slides and drawings of cajal. *Frontiers in neuroanatomy*, 4.
- [Golgi, 1873] Golgi, C. (1873). Sulla struttura della sostanza grigia del cervello. *Gazzetta Medica Italiana. Lombardia*, 33:244–246.
- [Grienberger and Konnerth, 2012] Grienberger, C. and Konnerth, A. (2012). Imaging calcium in neurons. *Neuron*, 73(5):862–885.
- [Hassabis et al., 2017] Hassabis, D., Kumaran, D., Summerfield, C., and Botvinick, M. (2017). Neuroscience-inspired artificial intelligence. *Neuron*, 95(2):245–258.
- [Hayworth et al., 2006] Hayworth, K., Kasthuri, N., Schalek, R., and Lichtman, J. (2006). Automating the collection of ultrathin serial sections for large volume tomographic reconstructions. *Microscopy and Microanalysis*, 12(S02):86.
- [Hayworth et al., 2015] Hayworth, K. J., Xu, C. S., Lu, Z., Knott, G. W., Fetter, R. D., Tapia, J. C., Lichtman, J. W., and Hess, H. F. (2015). Ultrastructurally smooth thick partitioning and volume stitching for large-scale connectomics. *Nature methods*, 12(4):319–322.
- [Hell, 2007] Hell, S. W. (2007). Far-field optical nanoscopy. *Science*, 316(5828):1153–1158.
- [Hell and Wichmann, 1994] Hell, S. W. and Wichmann, J. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782.

- [Helmstaedter, 2013] Helmstaedter, M. (2013). Cellular-resolution connectomics: challenges of dense neural circuit reconstruction. *Nature methods*, 10(6):501–507.
- [Helmstaedter et al., 2008a] Helmstaedter, M., Briggman, K. L., and Denk, W. (2008a). 3d structural imaging of the brain with photons and electrons. *Current opinion in neurobiology*, 18:633–641.
- [Helmstaedter et al., 2011] Helmstaedter, M., Briggman, K. L., and Denk, W. (2011). High-accuracy neurite reconstruction for high-throughput neuroanatomy. *Nature neuroscience*, 14(8):1081–1088.
- [Helmstaedter et al., 2013] Helmstaedter, M., Briggman, K. L., Turaga, S. C., Jain, V., Seung, H. S., and Denk, W. (2013). Connectomic reconstruction of the inner plexiform layer in the mouse retina. *Nature*, 500:168–174.
- [Helmstaedter et al., 2007] Helmstaedter, M., de Kock, C., Feldmeyer, D., Bruno, R. M., and Sakmann, B. (2007). Reconstruction of an average cortical column in silico. *Brain Research Reviews*, 55(2):193–203.
- [Helmstaedter et al., 2008b] Helmstaedter, M., Staiger, J. F., Sakmann, B., and Feldmeyer, D. (2008b). Efficient recruitment of layer 2/3 interneurons by layer 4 input in single columns of rat somatosensory cortex. *Journal of Neuroscience*, 28(33):8273–8284.
- [Hennig and Denk, 2007] Hennig, P. and Denk, W. (2007). Point-spread functions for backscattered imaging in the scanning electron microscope. *Journal of Applied Physics*, 102(12):123101.
- [Herculano-Houzel et al., 2006] Herculano-Houzel, S., Mota, B., and Lent, R. (2006). Cellular scaling rules for rodent brains. *Proceedings of the National Academy of Sciences*, 103(32):12138–12143.
- [Hildebrand et al., 2017] Hildebrand, D. G. C., Cicconet, M., Torres, R. M., Choi, W., Quan, T. M., Moon, J., Wetzel, A. W., Champion, A. S., Graham, B. J., Randlett, O., et al. (2017). Whole-brain serial-section electron microscopy in larval zebrafish. *bioRxiv*, page 134882.
- [Hinton et al., 2012] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T. N., et al. (2012). Deep

- neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *IEEE Signal Processing Magazine*, 29(6):82–97.
- [Hodgkin and Huxley, 1952] Hodgkin, A. L. and Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of physiology*, 117(4):500–544.
- [Hodgkin et al., 1952] Hodgkin, A. L., Huxley, A. F., and Katz, B. (1952). Measurement of current-voltage relations in the membrane of the giant axon of loligo. *The Journal of physiology*, 116(4):424–448.
- [Hua et al., 2015] Hua, Y., Laserstein, P., and Helmstaedter, M. (2015). Large-volume en-bloc staining for electron microscopy-based connectomics. *Nature communications*, 6:7923.
- [Hubel and Wiesel, 1959] Hubel, D. H. and Wiesel, T. N. (1959). Receptive fields of single neurones in the cat’s striate cortex. *The Journal of physiology*, 148(3):574–591.
- [Jain, 2017] Jain, V. (2017). Adversarial image alignment and interpolation. *arXiv preprint arXiv:1707.00067*.
- [Jain et al., 2010] Jain, V., Bollmann, B., Richardson, M., Berger, D. R., Helmstaedter, M. N., Briggman, K. L., Denk, W., Bowden, J. B., Mendenhall, J. M., Abraham, W. C., et al. (2010). Boundary learning by optimization with topological constraints. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 2488–2495. IEEE.
- [Januszewski et al., 2016] Januszewski, M., Maitin-Shepard, J., Li, P., Kornfeld, J., Denk, W., and Jain, V. (2016). Flood-filling networks. *arXiv preprint arXiv:1611.00421*.
- [Jiang et al., 2015] Jiang, X., Shen, S., Cadwell, C. R., Berens, P., Sinz, F., Ecker, A. S., Patel, S., and Tolias, A. S. (2015). Principles of connectivity among morphologically defined cell types in adult neocortex. *Science*, 350(6264):aac9462.
- [Jiang et al., 2003] Jiang, Y., Lee, A., Chen, J., Ruta, V., Cadene, M., Chait, B. T., and MacKinnon, R. (2003). X-ray structure of a voltage-dependent k<sup>+</sup> channel. *Nature*, 423:33–41.



- [Jones et al., 1952] Jones, W. H. S., Withington, E. T., et al. (1952). *Hippocrates: with an English translation*, volume 2. Harvard University Press.
- [Kasthuri et al., 2015] Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T. R., et al. (2015). Saturated reconstruction of a volume of neocortex. *Cell*, 162(3):648–661.
- [Kelley, 1960] Kelley, H. J. (1960). Gradient theory of optimal flight paths. *Ars Journal*, 30(10):947–954.
- [Kim et al., 2014] Kim, J. S., Greene, M. J., Zlateski, A., Lee, K., Richardson, M., Turaga, S. C., Purcaro, M., Balkam, M., Robinson, A., Behabadi, B. F., Campos, M., Denk, W., Seung, H. S., and EyeWriters (2014). Space-time wiring specificity supports direction selectivity in the retina. *Nature*, 509:331–336.
- [Knott et al., 2008] Knott, G., Marchman, H., Wall, D., and Lich, B. (2008). Serial section scanning electron microscopy of adult brain tissue using focused ion beam milling. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 28:2959–2964.
- [Ko et al., 2011] Ko, H., Hofer, S. B., Pichler, B., Buchanan, K., Sjöström, P. J., and Mrsic-Flogel, T. D. (2011). Functional specificity of local synaptic connections in neocortical networks. *Nature*, 473(7345):87.
- [Koelbl et al., 2013] Koelbl, C., Helmstaedter, M., Lübke, J., and Feldmeyer, D. (2013). A barrel-related interneuron in layer 4 of rat somatosensory cortex with a high intrabarrel connectivity. *Cerebral cortex*, 25(3):713–725.
- [Kreshuk et al., 2011] Kreshuk, A., Straehle, C. N., Sommer, C., Koethe, U., Cantoni, M., Knott, G., and Hamprecht, F. A. (2011). Automated detection and segmentation of synaptic contacts in nearly isotropic serial electron microscopy images. *PloS one*, 6(10):e24899.
- [Krizhevsky et al., 2012] Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105.
- [LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.

- [LeCun et al., 1989] LeCun, Y., Boser, B., Denker, J. S., Henderson, D., Howard, R. E., Hubbard, W., and Jackel, L. D. (1989). Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1(4):541–551.
- [Lee et al., 2016] Lee, W.-C. A., Bonin, V., Reed, M., Graham, B. J., Hood, G., Glattfelder, K., and Reid, R. C. (2016). Anatomy and function of an excitatory network in the visual cortex. *Nature*, 532:370–374.
- [Margrie et al., 2002] Margrie, T. W., Brecht, M., and Sakmann, B. (2002). In vivo, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Archiv European Journal of Physiology*, 444(4):491–498.
- [Markram, 2006] Markram, H. (2006). The blue brain project. *Nature reviews. Neuroscience*, 7(2):153.
- [Markram et al., 1997a] Markram, H., Lübke, J., Frotscher, M., Roth, A., and Sakmann, B. (1997a). Physiology and anatomy of synaptic connections between thick tufted pyramidal neurones in the developing rat neocortex. *The Journal of physiology*, 500(2):409–440.
- [Markram et al., 1997b] Markram, H., Lübke, J., Frotscher, M., and Sakmann, B. (1997b). Regulation of synaptic efficacy by coincidence of postsynaptic aps and epsps. *Science*, 275(5297):213–215.
- [Markram et al., 2015] Markram, H., Muller, E., Ramaswamy, S., Reimann, M. W., Abdellah, M., Sanchez, C. A., Ailamaki, A., Alonso-Nanclares, L., Antille, N., Arsever, S., et al. (2015). Reconstruction and simulation of neocortical microcircuitry. *Cell*, 163(2):456–492.
- [Marmont, 1949] Marmont, G. (1949). Studies on the axon membrane. i. a new method. *Journal of cellular physiology*, 34(3):351–382.
- [Mauss et al., 2017] Mauss, A. S., Vlasits, A., Borst, A., and Feller, M. (2017). Visual circuits for direction selectivity. *Annual Review of Neuroscience*, (0).
- [McCulloch and Pitts, 1943] McCulloch, W. S. and Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, 5(4):115–133.

- [Merriam-Webster, 2017] Merriam-Webster (24 July 2017). *Merriam-Webster Online dictionary*. Merriam-Webster. Consciousness.
- [Meyer et al., 2011] Meyer, H. S., Schwarz, D., Wimmer, V. C., Schmitt, A. C., Kerr, J. N., Sakmann, B., and Helmstaedter, M. (2011). Inhibitory interneurons in a cortical column form hot zones of inhibition in layers 2 and 5a. *Proceedings of the National Academy of Sciences*, 108(40):16807–16812.
- [Meyer et al., 2010a] Meyer, H. S., Wimmer, V. C., Hemberger, M., Bruno, R. M., de Kock, C. P., Frick, A., Sakmann, B., and Helmstaedter, M. (2010a). Cell type-specific thalamic innervation in a column of rat vibrissal cortex. *Cerebral cortex*, 20(10):2287–2303.
- [Meyer et al., 2010b] Meyer, H. S., Wimmer, V. C., Oberlaender, M., De Kock, C. P., Sakmann, B., and Helmstaedter, M. (2010b). Number and laminar distribution of neurons in a thalamocortical projection column of rat vibrissal cortex. *Cerebral cortex*, 20(10):2277–2286.
- [Mikula, 2016] Mikula, S. (2016). Progress towards mammalian whole-brain cellular connectomics. *Frontiers in neuroanatomy*, 10.
- [Mikula and Denk, 2015] Mikula, S. and Denk, W. (2015). High-resolution whole-brain staining for electron microscopic circuit reconstruction. *Nature methods*, 12(6):541–546.
- [Minsky, 1961] Minsky, M. (1961). Microscopy apparatus us patent 3013467. *USP Office, Ed. US*.
- [Miyawaki et al., 1997] Miyawaki, A., Llopis, J., Heim, R., McCaffery, J. M., et al. (1997). Fluorescent indicators for  $Ca^{2+}$  based on green fluorescent proteins and calmodulin. *Nature*, 388(6645):882.
- [Nakai et al., 2001] Nakai, J., Ohkura, M., and Imoto, K. (2001). A high signal-to-noise  $Ca^{2+}$  probe composed of a single green fluorescent protein. *Nature biotechnology*, 19(2):137.
- [Narayanan et al., 2015] Narayanan, R. T., Egger, R., Johnson, A. S., Mansvelder, H. D., Sakmann, B., De Kock, C. P., and Oberlaender, M. (2015). Beyond columnar organization: cell type-and target layer-specific principles of horizontal axon projection patterns in rat vibrissal cortex. *Cerebral cortex*, 25(11):4450–4468.

- [Neher and Sakmann, 1976] Neher, E. and Sakmann, B. (1976). Single-channel currents recorded from membrane of denervated frog muscle fibres. *Nature*, 260(5554):799–802.
- [Pallotto et al., 2015] Pallotto, M., Watkins, P. V., Fubara, B., Singer, J. H., and Briggman, K. L. (2015). Extracellular space preservation aids the connectomic analysis of neural circuits. *Elife*, 4(4).
- [Peters and Feldman, 1976] Peters, A. and Feldman, M. L. (1976). The projection of the lateral geniculate nucleus to area 17 of the rat cerebral cortex. i. general description. *Journal of Neurocytology*, 5(1):63–84.
- [Preibisch et al., 2009] Preibisch, S., Saalfeld, S., and Tomancak, P. (2009). Globally optimal stitching of tiled 3d microscopic image acquisitions. *Bioinformatics*, 25(11):1463–1465.
- [Prioreschi, 1991] Prioreschi, P. (1991). Possible reasons for neolithic skull trephining. *Perspectives in biology and medicine*, 34:296–303.
- [Rand, 1971] Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical association*, 66(336):846–850.
- [Renshaw et al., 1940] Renshaw, B., Forbes, A., and Morison, B. (1940). Activity of isocortex and hippocampus: electrical studies with micro-electrodes. *Journal of Neurophysiology*, 3(1):74–105.
- [Roncal et al., 2014] Roncal, W. G., Pekala, M., Kaynig-Fittkau, V., Kleissas, D. M., Vogelstein, J. T., Pfister, H., Burns, R., Vogelstein, R. J., Chevillet, M. A., and Hager, G. D. (2014). Vesicle: Volumetric evaluation of synaptic interfaces using computer vision at large scale. *arXiv preprint arXiv:1403.3724*.
- [Rosenblatt, 1958] Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.
- [Rust et al., 2006] Rust, M. J., Bates, M., and Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm). *Nature methods*, 3(10):793–796.
- [Sakmann and Neher, 1984] Sakmann, B. and Neher, E. (1984). Patch clamp techniques for studying ionic channels in excitable membranes. *Annual review of physiology*, 46(1):455–472.

- [Schmidt et al., 2017] Schmidt, H., Gour, A., Straehle, J., Boergens, K. M., Brecht, M., and Helmstaedter, M. (2017). Axonal synapse sorting in medial entorhinal cortex. *Nature*, advance online publication.
- [Schneider-Mizell et al., 2016] Schneider-Mizell, C. M., Gerhard, S., Longair, M., Kazimiers, T., Li, F., Zwart, M. F., Champion, A., Midgley, F. M., Fetter, R. D., Saalfeld, S., et al. (2016). Quantitative neuroanatomy for connectomics in drosophila. *Elife*, 5:e12059.
- [Sporns et al., 2005] Sporns, O., Tononi, G., and Ktter, R. (2005). The human connectome: A structural description of the human brain. *PLoS computational biology*, 1:e42.
- [Staffler et al., 2017] Staffler, B., Berning, M., Boergens, K. M., Gour, A., van der Smagt, P., and Helmstaedter, M. (2017). SynEM: Automated synapse detection for connectomics. *eLife*, 6:e26414.
- [Südhof, 1995] Südhof, T. C. (1995). The synaptic vesicle cycle: a cascade of protein protein interactions. *Nature*, 375(6533):645–653.
- [Südhof and Rothman, 2009] Südhof, T. C. and Rothman, J. E. (2009). Membrane fusion: grappling with snare and sm proteins. *Science*, 323(5913):474–477.
- [Takemura et al., 2017] Takemura, S.-Y., Aso, Y., Hige, T., Wong, A., Lu, Z., Xu, C. S., Rivlin, P. K., Hess, H. F., Zhao, T., Parag, T., Berg, S., Huang, G., Katz, W., Olbris, D. J., Plaza, S., Umayam, L., Aniceto, R., Chang, L.-A., Lauchie, S., Ogundeyi, O., Ordish, C., Shinomiya, A., Sigmund, C., Takemura, S., Tran, J., Turner, G. C., Rubin, G. M., and Scheffer, L. K. (2017). A connectome of a learning and memory center in the adult drosophila brain. *eLife*, 6.
- [Takemura et al., 2013] Takemura, S.-y., Bharioke, A., Lu, Z., Nern, A., Vitaladevuni, S., Rivlin, P. K., Katz, W. T., Olbris, D. J., Plaza, S. M., Winston, P., Zhao, T., Horne, J. A., Fetter, R. D., Takemura, S., Blazek, K., Chang, L.-A., Ogundeyi, O., Saunders, M. A., Shapiro, V., Sigmund, C., Rubin, G. M., Scheffer, L. K., Meinertzhagen, I. A., and Chklovskii, D. B. (2013). A visual motion detection circuit suggested by drosophila connectomics. *Nature*, 500:175–181.
- [Takemura et al., 2015] Takemura, S.-y., Xu, C. S., Lu, Z., Rivlin, P. K., Parag, T., Olbris, D. J., Plaza, S., Zhao, T., Katz, W. T., Umayam, L., Weaver, C., Hess, H. F.,

- Horne, J. A., Nunez-Iglesias, J., Aniceto, R., Chang, L.-A., Lauchie, S., Nasca, A., Ogundeyi, O., Sigmund, C., Takemura, S., Tran, J., Langille, C., Le Lacheur, K., McLin, S., Shinomiya, A., Chklovskii, D. B., Meinertzhagen, I. A., and Scheffer, L. K. (2015). Synaptic circuits and their variations within different columns in the visual system of drosophila. *Proceedings of the National Academy of Sciences of the United States of America*, 112:13711–13716.
- [Trachtenberg et al., 2002] Trachtenberg, J. T., Chen, B. E., Knott, G. W., Feng, G., et al. (2002). Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. *Nature*, 420(6917):788.
- [Tsien, 1983] Tsien, R. (1983). Calcium channels in excitable cell membranes. *Annual review of physiology*, 45(1):341–358.
- [Turaga et al., 2009] Turaga, S., Briggman, K., Denk, W., Seung, S., and Helmstaedter, M. (2009). Maximin affinity learning of image segmentation. In *Advances in Neural Information Processing Systems*, pages 1865–1873.
- [Turaga et al., 2010] Turaga, S. C., Murray, J. F., Jain, V., Roth, F., Helmstaedter, M., Briggman, K., Denk, W., and Seung, H. S. (2010). Convolutional networks can learn to generate affinity graphs for image segmentation. *Neural computation*, 22(2):511–538.
- [Turing, 1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, 59(236):433–460.
- [Varshney et al., 2011] Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H., and Chklovskii, D. B. (2011). Structural properties of the caenorhabditis elegans neuronal network. *PLoS computational biology*, 7(2):e1001066.
- [Vincent, 1913] Vincent, S. (1913). The tactile hair of the white rat. *Journal of comparative neurology*, 23(1):1–34.
- [Vincent, 1912] Vincent, S. B. (1912). *The functions of the vibrissae in the behavior of the white rat...*, volume 1. University of Chicago.
- [Volta, 1800] Volta, A. (1800). On the electricity excited by the mere contact of conducting substances of different kinds. in a letter from mr. alexander volta, frs professor of natural philosophy in the university of pavia, to the rt. hon. sir joseph banks, bart. kbprs. *Philosophical transactions of the Royal Society of London*, pages 403–431.

- [Waite et al., 1991] Waite, P. M., Marotte, L. R., and Mark, R. F. (1991). Development of whisker representation in the cortex of the tammar wallaby *macropus eugenii*. *Brain research. Developmental brain research*, 58:35–41.
- [Wanner et al., 2016] Wanner, A. A., Genoud, C., Masudi, T., Siksou, L., and Friedrich, R. W. (2016). Dense em-based reconstruction of the interglomerular projectome in the zebrafish olfactory bulb. *Nature neuroscience*, 19(6):816–825.
- [Ward et al., 1975] Ward, S., Thomson, N., White, J. G., and Brenner, S. (1975). Electron microscopical reconstruction of the anterior sensory anatomy of the nematode *caenorhabditis elegans*. *The Journal of comparative neurology*, 160:313–337.
- [Wernicke, 1874] Wernicke, C. (1874). *Der aphasische Symptomencomplex: eine psychologische Studie auf anatomischer Basis*. Cohn.
- [White et al., 1986] White, J. G., Southgate, E., Thomson, J. N., and Brenner, S. (1986). The structure of the nervous system of the nematode *caenorhabditis elegans*. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 314:1–340.
- [Willingham and Rutherford, 1984] Willingham, M. C. and Rutherford, A. V. (1984). The use of osmium-thiocarbohydrazide-osmium (oto) and ferrocyanide-reduced osmium methods to enhance membrane contrast and preservation in cultured cells. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 32:455–460.
- [Wimmer et al., 2010] Wimmer, V. C., Bruno, R. M., De Kock, C. P., Kuner, T., and Sakmann, B. (2010). Dimensions of a projection column and architecture of vpm and pom axons in rat vibrissal cortex. *Cerebral cortex*, 20(10):2265–2276.
- [Wong-Riley, 1989] Wong-Riley, M. T. (1989). Cytochrome oxidase: an endogenous metabolic marker for neuronal activity. *Trends in neurosciences*, 12(3):94–101.
- [Woolsey and Van der Loos, 1970] Woolsey, T. A. and Van der Loos, H. (1970). The structural organization of layer iv in the somatosensory region (si) of mouse cerebral cortex: the description of a cortical field composed of discrete cytoarchitectonic units. *Brain research*, 17(2):205–242.

- [Woolsey et al., 1975] Woolsey, T. A., Welker, C., and Schwartz, R. H. (1975). Comparative anatomical studies of the sml face cortex with special reference to the occurrence of barrels in layer iv. *Journal of Comparative Neurology*, 164(1):79–94.
- [Zheng et al., 2017] Zheng, Z., Lauritzen, J. S., Perlman, E., Robinson, C. G., Nichols, M., Milkie, D., Torrens, O., Price, J., Fisher, C. B., Sharifi, N., et al. (2017). A complete electron microscopy volume of the brain of adult drosophila melanogaster. *bioRxiv*, page 140905.



## 9 Acknowledgements

First and foremost I want to thank Patricia for her patience and consideration during these busy years during which I sometimes was not able to allocate the time to our relationship that I would have liked to.

For his large contribution to all projects summarized here, I want to thank Moritz. He has been the driving influence of approaches to explore and which methods to use and a major motivating factor and coordinator in all projects presented here.

Furthermore, this work would not have been possible without an intense collaboration with Kevin who acquired and aligned the dataset used in the method developments presented in this thesis. He was a shared first author in [Boergens et al., 2017], provided input and interesting discussion in all projects. He also has become a good friend.

I also want to thank Benedikt, who is the first author of [Staffler et al., 2017] and who has contributed a lot to algorithmic development and improvement of the shared code with specific focus on interface classification and data structures.

Alessandro joined the lab during the later phase of my PhD. I want to thank him for being a great discussion partner. He has contributed to the method developments described in the outlook in major ways.

Jakob for being a good friend and my main discussion partner and source of information for all topics relating to biology, anatomy or medicine.

All other members in the lab I want to thank whole heartily for generating a nice atmosphere and interesting scientific environment.

My friends and family I want to thank for not abandoning me even though I did not have any time!

# 10 Curriculum Vitae

## Manuel Berning

**Current address:**

Roedelheimer Landstr. 68  
60487 Frankfurt am Main  
Germany

**Date of birth:** June 17th, 1985

**Place of birth:** Mettingen, Germany

**Phone:** +49 173 5243322

**Email:** manuel.mb.berning@gmail.com

### Education

|              |  |
|--------------|--|
| 2011–present | Doctoral research at the Max-Planck Institute of Neurobiology in Martinsried and the Max-Planck Institute for Brain Research in Frankfurt working on "Computational methods in connectomics"   |
| 2010–2011    | Diploma thesis "Simulation of the neuronal circuitry in the glomerular layer of the olfactory bulb with special emphasis on phase shifts of action potential firing with respect to the respiratory cycle" at the Max-Planck Institute of Medical Research, Heidelberg |
| 2005–2011    | M.Sc. (Diplom) in physics with minors in computer science and biophysics at Karl-Ruprecht-University, Heidelberg   |
| 1995–2004    | High school diploma (Abitur) at Norbert Gymnasium, Knechtsteden  |
| 2002–2004    | Education as "Junior Manager" at the Junior Management School, Knechtsteden  |
| 2001–2002    | Student exchange to Great Valley High School in Pennsylvania, USA  |

## Publications

- |      |   |
|------|---|
| 2017 | B. Staffler, <b>M. Berning</b> , K.M. Boergens, A. Gour, P. van der Smagt, M. Helmstaedter. SynEM: Automated synapse detection for connectomics. <i>eLife</i> 2017;6:e26414   |
| 2017 | K.M. Boergens*, <b>M. Berning*</b> , T. Bocklisch, D. Braeunlein, F. Drawitsch, J. Frohnhofen, T. Herold, P. Otto, N. Rzepka, T. Werkmeister, D. Werner, G. Wiese, H. Wissler, M. Helmstaedter. webKnossos: efficient online 3D data annotation for connectomics. <i>nature methods</i> 14(7), 691-694 (* = shared first authors) |
| 2015 | <b>M. Berning</b> , K.M. Boergens, M. Helmstaedter. SegEM: efficient image analysis for high-resolution connectomics. <i>Neuron</i> , 87(6), 1193-1206.   |
| 2012 | I. Fukunaga, <b>M. Berning</b> , M. Kollo, A. Schmaltz, A.T. Schaefer. Two distinct channels of olfactory bulb output. <i>Neuron</i> , 75(2), 320-329.  |

## Experiences

|             |  |
|-------------|--|
| Languages   | English (fluent) and German (native)   |
| Programming | Matlab, Python, C/C++, FORTRAN, PHP, Javascript, Delphi, Visual Basic, Basic |
| Tools       | Linux, Bash, Blender, NEURON, Adobe CS, Office                               |
| Other       | 2004/05 Community service as a paramedic for the German Red Cross            |

## Poster presentations, lectures and summer schools

|      |  |
|------|--|
| 2017 | Lecture and tutorial session as part of the FENS Cajal Course "Interacting with neural circuits", Lisbon, Portugal   |
| 2017 | Invited speaker at the Connectomics conference, Berlin, Germany  |
| 2016 | Poster presentation at the Connectomics conference, Ashburn, United States   |
| 2016 | Janelia DVID (Distributed, Versioned, Image-oriented Dataservice) Workshop and Hackathon, Ashburn, United States   |
| 2015 | Invited speaker at the Connectome Workbench, Cambridge, England  |
| 2015 | Invited speaker at the Connectomics conference, Berlin, Germany  |
| 2015 | Janelia EM Connectomics Analysis Workshop and Hackathon, Ashburn, United States  |
| 2015 | Society for Neuroscience Conference poster presentation, Chicago, United States  |
| 2014 | CapoCaccia Cognitive Neuromorphic Engineering Workshop, Alghero, Italy   |
| 2014 | Society for Neuroscience Conference poster presentation, Washington, United States   |
| 2013 | Society for Neuroscience Conference poster presentation, San Diego, United States  |
| 2012 | Host for Neural Information Processing Systems (NIPS) workshop "Connectomics: Opportunities and Challenges for Machine Learning", Lake Tahoe, United States                |
| 2012 | Machine Learning Summer School (MLSS), La Palma, Spain   |
| 2012 | Society for Neuroscience Conference poster presentation, New Orleans, United States  |
| 2012 | Talks on "Machine Learning of image segmentation for the dense reconstruction of neuronal circuits" at joint MPG and DELL info events, Halle, Frankfurt, Aschheim, Germany |